Argument Schemes and Critical Questions for Ethical Reasoning

E.Bezou-Vrakatseli, O. Cocarascu and S. Modgil

Abstract

Atkinson's practical reasoning scheme (PRS) and its associated sixteen critical questions have been widely deployed to support the construction of arguments and counter-arguments that inform both computational and human decision making. PRS evaluates actions in terms of the values they promote; the latter being typically articulated in generic terms (e.g., fairness, justice, etc.). However, amidst ongoing concerns about how to align computational decisions with humans' ethically based preferences, there is a need to to develop more specialised schemes and critical questions that are tailored to a more ecumenical accounting of the variety of ethical theories that have been developed within moral philosophy. In this paper we therefore propose argument schemes and critical (AS&CQ)questions that capture key features of consequentialist, deontic and virtue-ethics based theories. We illustrate, using a number of ethically salient debates, how this broader and more specialised AS&C support the dialectical exchange of arguments within these debates, and how these AS&CQ can be computationally operationalised in $ASPIC+$ criteria that

1 Introduction

TBD

2 Methodology

The development of the new argument schemes was based on varying the *argument from action* scheme to accommodate the four major ethical theories. This adaptation involved expanding the scheme's multi-variable approach by introducing theory-specific variables that capture the distinctive justifications and inference rules of each framework. These variables not only define the structural components of each scheme but also identify the aspects most susceptible to critical challenges. The schemes' critical questions were, thus, also crafted by drawing on both the descriptive insights of Section **??** and established philosophical critiques of these theories.

Similarly to *argument from action*, the new schemes incorporate multiple variables to represent circumstances, the action in question, and the main components of an ethical theory. This way, competing (or supporting) arguments can be represented transparently, making it possible to identify the sources of conflicting conclusions and enable reasoning about them. Consider an *argument from action*, asserting that in the given circumstances R, we should perform action A, because it promotes value V. An opposing argument might contend that action A demotes a different value V', leading to mutually attacking arguments that appeal to different ethical considerations. The result is a pair of mutually attacking arguments grounded in distinct ethical considerations, exemplifying a typical ethical dilemma in which competing values must be weighed. The scheme thus helps to reveal the core of the conflict.

As discussed in Section **??**, such arguments can be formalised in *ASPIC*$^+$ and subsequently represented within a Dung framework [**?**] to determine the set of justified (i.e. winning) arguments. Moreover, in cases of symmetrical conflict such as this, *ASPIC*$^+$ allows predefined preferences over premises to be specified, thereby providing a principled means of prioritising one argument over another to resolve the dispute.

Analogously, the new schemes developed in this Section allow to treat entities, such as virtues, rules, rights, and evaluative metrics of consequences, as features that can be compared, prioritised, and challenged. This supports the formalisation of reasoning about ethical preferences. To orient the reader, I first revisit the original *argument from action* in Box **??**, followed by its corresponding critical questions, before presenting the proposed variations for ethical theories.

Box 1 Argument from Action Scheme

In the current circumstances R
We should perform action A
Which will result in new circumstances S
Which will realise goal G
Which will promote value V.

CQ1: Are the believed circumstances true?
CQ2: Assuming the circumstances, does the action have the stated consequences?
CQ3: Assuming the circumstances and that the action has the stated consequences, will the action bring about the desired goal?
CQ4: Does the goal realise the value stated?
CQ5: Are there alternative ways of realising the same consequences?
CQ6: Are there alternative ways of realising the same goal?
CQ7: Are there alternative ways of promoting the same value?
CQ8: Does doing the action have a side effect which demotes the value?
CQ9: Does doing the action have a side effect which demotes some other value?
CQ10: Does doing the action promote some other value?
CQ11: Does doing the action preclude some other action which would promote some other value?
CQ12: Are the circumstances as described possible?
CQ13: Is the action possible?
CQ14: Are the consequences as described possible?
CQ15: Can the desired goal be realised?
CQ16: Is the value indeed a legitimate value?

To develop the variations on this scheme, I followed a structured approach to capture the subtleties of ethical reasoning.

1. Study of *Kialo*[1] debates

   The analysis of debates on *Kialo* detailed in Section **??** was instrumental in understanding how frequently arguments aligned with the schemes appealing to ethical theories. It highlighted the

[1] https://www.kialo.com

necessity of these schemes, suggesting that they could enhance the quality of debates by providing clearer guidance on constructing arguments. This exploration also allowed to identify the critical questions that frequently emerge in ethical debates. By examining real-world examples, I was able to pinpoint recurring patterns that these new schemes needed to address, ensuring they were both relevant and practical for actual discourse.

2. Adaptation of the *argument from action* scheme

   I modified the original *argument from action* to accommodate the main elements of the four main ethical frameworks. This required modifying the phrasing and, in some cases, the structure of the scheme to align with the foundational principles of each ethical theory.

3. Introduction of new variables

   I formalised the new schemes in a manner analogous to the original *argument from action*, introducing variables that represent the key factors involved in each ethical framework. These variables not only define the structure of each scheme but also highlight the aspects most susceptible to challenge. Identifying these critical points helps inform, in return, the development of the corresponding critical questions.

4. Critical questions (CQs)

   Regarding the new critical questions, I retained the relevant questions from the original *argument from action* and incorporated additional ones specific to each scheme. This enhancement was informed both by insights from the study of debates on *Kialo* and by established criticisms of the ethical theories themselves, as developed within moral philosophy. The critical questions serve various purposes:

   - Elicit supporting arguments: prompt the interlocutor to supply additional reasons that strengthen the original argument.
   - Elicit attacking arguments: invite objections or counter-evidence that challenge the original argument.
   - Extend a premise: request the completion or clarification of an implicit or abbreviated premise (e.g., filling an enthymeme).
   - Elicit further information: gather extra contextual details (e.g., for the purpose of later exposing inconsistencies or tensions within the argument).

5. Prescriptive study of formal debates

   As a last step, I checked whether the new argument schemes and critical questions represent ethical argumentation accurately by examining how well they capture arguments used in formal debates. Thus, I fine-tuned the schemes on formal debates to ensure they not only captured the essence of ethical reasoning in everyday discourse, but also provided a guide to how such debates should ideally unfold.

3 Action Scheme from Virtue Ethics

Virtue Ethics emphasises the role of character and virtue in moral philosophy. A virtue is an excellent trait of character and having a virtue means being a certain type of person with a specific mindset. A crucial component of this mindset is fully embracing a specific set of considerations as reasons for action. In Virtue

Virtue Ethics	
Focus	Character / Virtues
Question	"Who should I be?"
Founder	Aristotle
Core	Cultivation of virtues
Strengths	Holistic moral development
Weaknesses	Lacks decision rules

Deontology	
Focus	Duty / Rules
Question	"What must I do?"
Founder	I. Kant
Core	Universal moral laws
Strengths	Clear moral rules
Weaknesses	Can be rigid

Consequentialism	
Focus	Consequences
Question	"What has the best outcome?"
Founder	J. Bentham, J. S. Mill
Core	Maximise happiness
Strengths	Practical, outcome-driven
Weaknesses	May justify harmful means

Rights-Based Ethics	
Focus	Rights
Question	"What rights must be respected?"
Founder	J. Locke, R. Dworkin
Core	Respect human rights
Strengths	Protects individual freedoms
Weaknesses	Conflicting rights hard to resolve

Table 1: Comparison of the four main normative ethical theories.

Ethics, what makes a person moral/virtuous is practicing good habits, such as courage, humility, bravery, respect, honesty, generosity,etc. Virtue ethicists see virtues and vices as foundational for virtue ethical theories, and other normative notions are grounded in them [**?**]. In virtue-based reasoning, arguments often appeal to character traits such as *being responsible*. Consider the following argument from the *Kialo* debate *Pro-Life vs Pro-Choice: Should Abortion be Legal?*, where the virtue of responsibility is invoked.

> A : "A person who chooses to engage in sexual activity should accept the responsibility of the potential consequences, including pregnancy."

Action Scheme from Virtue Ethics is developed to model such arguments, which are motivated by and appeal to Virtue Ethics. For this, I adapt the *argument from action* scheme [**?**] as follows. The scheme is presented in Box **??**.

- I introduce the concept of a virtuous agent VA.
- I establish the moral virtues MV, according to which a VA acts.
- I set the motive for an action to be directly linked to MV.
- I link the action A, its effects E, and its goal G directly to the circumstances R. I put the goal G in brackets, as the expression of a goal G is not necessary for virtuous reasoning; however, it is relevant in the broader argumentative context, as interlocutors may introduce considerations about it (i.e., critical questions regarding the goal are relevant in the dialogue).

When formulating the scheme and its critical questions, I make use of the following terminology.

- Action A *expresses/represents/exemplifies* a moral virtue MV when it is in accordance with it.
  For instance, $A =$"Acknowledging to a friend that you made a mistake that affected them" exemplifies $MV = Honesty$.
- Action A *does not exemplify* a moral virtue MV if it is irrelevant to the virtue.
  For instance, $A =$"Going to work" does not exemplify $MV = Honesty$.

- Action A *is in conflict with* a moral virtue MV if it exemplifies the opposite of MV.
  For instance, A ="Deliberately falsifying information on a job application" is in conflict with $MV = Honesty$.[2]

Box 1 Action Scheme from Virtue Ethics

In the current circumstances R
We should perform action A
Which has effect E
{Which will realise goal G}
Since a virtuous agent VA would perform action A in circumstances R
Exemplifying virtue MV

The original example was an argument found 'in the wild' positing that one should accept the consequences of their actions, implicitly advocating for action A "one should carry through with the pregnancy". I now construct argument A_r, a refined example for action A to illustrate the virtue-based reasoning scheme more formally, where $R = Pregnancy$, $MV = Responsibility$, $A = StayPregnant$, $E = HaveBaby$, $G = PreserveLife$.

> A_r : "If a virtuous person were to become pregnant, they would act responsibly by carrying the pregnancy to term so that life is preserved. Therefore, when someone becomes pregnant, they ought to carry through with the pregnancy, which ultimately results in the birth of a child."

A representation of the scheme in *ASPIC*$^+$ is presented below.
f_1 : *We are in circumstances* R.
f_2 : *In* R*, action* A *has effect* E.
$\{f_3$: *In* R*, action* A *realises goal* $G.\}$
f_4 : *In* R*, a virtuous agent* VA *would perform action* A*, exemplifying virtue* MV.

d_v (R, VA, A, MV, E, G): $Circumstances(R)$, $ExemplifiesVirtue(R, A, VA, MV)$, $Effect(R, A, E)$, $Goal(R, A, G) \Rightarrow ShouldPerform(R, A)$

$$A_r : [f_1, f_2, \{f_3\}, f_4] \Rightarrow_{d_v} ShouldPerform(R, A)$$

3.1 Critical Questions

I additionally introduce a comprehensive set of critical questions for this scheme. First, I retain certain critical questions from the original *argument from action*. I slightly adjust some of those to the context of the theory (e.g., I significantly expand on CQ16, which interrogates the legitimacy of value V). Additionally, I introduce new critical questions derived from the main criticisms of Virtue Ethics. For each critical question, I include relevant examples and offer detailed explanations tailored to the specific aims of the framework. The questions are methodically grouped to enhance clarity and facilitate their deployment, purporting to guide the selection of the appropriate questions in order to effectively challenge and assess the acceptability of arguments.

[2]For virtue models that represent virtues and vices as opposites (e.g., [**?**]), the expression "is in conflict with MV" would be synonymous with "exemplifies the vice associated with MV"; in this case, the vice of dishonesty.

Group 1: Circumstances and Feasibility I maintain the original scheme's CQs inquiring about the feasibility and truthfulness in terms of both circumstances and proposed action. I first question the truthfulness of the circumstances, which assesses the argument's immediate applicability to known, factual contexts:

Are circumstances R true?

I then question the plausibility/possibility of the circumstances to facilitate a broader analysis, crucial in policy-making contexts and ethical debates where hypothetical scenarios are prevalent:

Are the circumstances R plausible/possible?

In general, the distinction between truthfulness and plausibility/possibility of circumstances is crucial when distinguishing practical ethics from meta-ethics. Practical ethics involves concrete, actionable ethical decision-making in current, real-world situations; thus, it depends significantly on the truthfulness of the circumstances. In contrast, meta-ethics involves examining the plausibility and coherence of ethical frameworks, including hypothetical scenarios that might inform ethical principles or policies. Policy debates, in particular, very often necessitate a broader examination that encompasses not only the factual, but also the potential. Questioning the overall plausibility (or even possibility) of the stated circumstances ensures that ethical reasoning and policymaking remain grounded in realistic and coherent possibilities, and avoid engaging in potentially irrational or unnecessary actions prompted by impossible or absurd scenarios.

The critical question regarding the possibility of circumstances also becomes useful when the truthfulness of specific circumstances is difficult or impossible to prove or reject directly. Consider, for instance, an argument advocating some action based on a religious prophecy, suggesting the allocation of substantial public resources to avert a predicted catastrophe. While directly disproving the truthfulness of the assumed, prophetic scenario might be challenging due to its inherently speculative or faith-based nature, one could nonetheless question, "Are these circumstances even possible?". Appealing to widely accepted scientific knowledge allows participants in the debate to dismiss implausible scenarios, even in the absence of direct proof of their truthfulness.

Moreover, this CQ also points to a higher, *dialectical* layer: beyond premise-level assessment, one may challenge the speaker's epistemic/experiential standing to assess responsibilities under R. In the running example, if a person is not pregnant, or is incapable of becoming pregnant, arguments about responsibilities tied to pregnancy are irrelevant to their personal context. This distinction matters especially in hypothetical discussions. Ethical debates frequently invoke counterfactuals (e.g., "What if you were pregnant?"), where R is not true but may be possible. Here the distinction is between agents for whom R is biologically or practically possible and those for whom it is not; many argue that one cannot meaningfully engage a hypothetical one could never realistically experience. This point recurs in abortion discourse, particularly in debates about legislative authority. A notable criticism is that many policymakers making decisions on abortion rights are individuals who are not (and have never been) able to experience pregnancy. Such debates involve an implicit critique regarding the relevance and legitimacy of perspectives from those whose lived experiences do not align with the circumstances under discussion.

Moreover, I introduce a question inquiring whether all relevant factors have been stated. Even if the previous two checks hold, and the circumstances are possible and true as stated, they may still fall short of providing a sufficiently informative or faithful representation of the situation. To address this, I add the following question:

Are circumstances R as stated complete?

For example, consider an argument against abortion claiming that "pregnant women should be responsible and carry to term", thus framed under the circumstances $R = Pregnancy$. If the actual situation is one

of $R = LifeThreateningPregnancy$, this framing omits a crucial condition and thus misrepresents the circumstances. Formally, we can model the base circumstance as $R =$ *Pregnancy* and the more specific circumstances as $R' = LifeThreateningPregnancy$. I also introduce additional predicate $LT(R, W)$ indicating that the pregnancy is life-threatening for the woman W.

$$C(R') \equiv \big(C(R) \wedge LT(R, W)\big).$$

This separates the *base circumstance* (pregnancy) from the *qualifying condition* (life-threatening). A counter argument could, then, attack premise f_1 by saying "You are misrepresenting the circumstances, the actual circumstances entail dangers for the pregnant person".[3]

Next, I investigate the possibility of the proposed action in the stated circumstances.

Is action A possible in circumstances R?

In the example argument, the feasibility question would be: "Is carrying to term possible for this person?". For instance, if continuing the pregnancy is medically impossible, then the proposed action (i.e., remaining pregnant) is not a valid option. This illustrates an important point: assessing feasibility is dependent on circumstances R. I therefore refine the original CQ for *arguments from action* to highlight the dependency of the feasibility of the action to circumstances R. In this example, this ties in with the previous point about circumstances and how accurately R has been defined and delimited, but this is not necessarily always the case. Consider a different example: someone argues that abortion should be illegal and, for teenagers, parents should monitor their children's every move as "the responsible thing to do". This proposed action is effectively impossible.

Thus, the CQ generates an attack on the argument's assumption that the person can perform the proposed action A. If we perceive the feasibility of A as an implicit condition of the defeasible rule d_v, then ¬Possible(R, A) attacks the rule's applicability. For example, the defeasible rule can be rephrased as "If action A has effect E (and no obstacles), so it realises G and exemplifies MV, then do A". An argument that states some obstacle, rendering A impossible, undercuts the application of that rule. The undercutter that responds to this CQ can, therefore, be represented as:

$u_v(A)$: $\neg\, Possible(R, A) \Rightarrow \neg\, d_v$ (R, VA, A, MV, E, G)

Group 2: Action's Effects and Goal Once circumstances R and the possibility of the action A have been established, I then assess whether A produces the intended consequences. The second group of CQs draws again directly from the *argument from action* and investigates the link between action A and its outcomes; i.e., effect E and goal G. These questions are the following.

Does action A have effect E in R?

Will action A realise goal G in R?

As mentioned earlier, even though a Virtue Ethics argument might not make direct reference to the outcomes of the proposed action—as opposed to a consequentialist scheme (see Section **??**)—the action's effects are decisive. The CQs in this group point to potential undermining attacks that challenge the scheme's premises $f_2 : Effect(R, A, E)$ and $f_3 : Goal(R, A, G)$.

[3]Depending on the framing, the "life-threatening" aspect could alternatively be placed under Group 2 (about the action's effects) rather than under circumstances. The distinction is often context-sensitive; many properties can be described either as part of the initial situation (R) or as an expected outcome (E).

Consider, for example, a case where Region X has suffered severe flooding, displacing thousands of people (R). One could make an argument instantiating *Action Scheme from Virtue Ethics* for action $A =$"Donate to charity Z's emergency appeal", exemplifying the virtue $MV = Benevolence$. A counter-argument, responding to the former CQ could be, "In Region X, all bank operations have been frozen after the floods and Charity Z's last two emergency appeals in X failed to deliver cash transfers". This argument attacks the premise that in the circumstances of these floods, the proposed action would have anticipated effect $E =$ "Funds reach displaced families", an outcome central to the action's exemplification of the cited virtue. So, in circumstances R, A does not have effect E.

If we assume the goal of the proposed action to be $G =$"Alleviate immediate suffering", if A does not have E, it (most likely) fails to realise G as well. But even in cases where the funds might reach the intended recipients, inefficiencies could undermine the impact required to achieve the goal (e.g., maybe past evidence indicates that, even when resources arrive, they are misallocated or misused).

The last question of this group addresses potential side effects E' of action A that might be in conflict with MV:

> *In R, does performing action A have a side effect E' that is in conflict with the moral virtue MV?.*

For instance, returning to the abortion-related arguments, consider the argument A_s: "Choosing to raise a child when one is not ready is an irresponsible decision". This argument focuses on a later effect of the action (i.e., the bringing up of the child), which contradicts MV. This brings us to the next group of CQs that focus on the link between action A and virtue MV.

Group 3: Link between Action and Virtue The next group of critical questions derives from challenging whether the proposed action expresses the appealed virtue. I first introduce the following CQ:

> *(How) does action A exemplify the virtues MV identified as relevant in R?.*

For example, in the original example, one might ask, "How is carrying on an unwanted pregnancy exemplifying responsibility?". This CQ targets the connection between the action and the virtue, pointing at an undermining attack, as it challenges premise f_4. It is intended as a broad inquiry to elicit further information, prompting the arguer to articulate the effects or goals that justify the perceived virtue. This is especially useful when the argument is not fully developed (like the idealised argument A_r) but, rather, is presented in a more implicit form (as the original argument A from *Kialo* that is an enthymeme). In this case, this CQ helps to flesh out the rest of the enthymeme. To further scrutinise this connection, I propose additional sub-questions.

The first investigates context sensitivity and how actions can differentially express virtues depending on the circumstances:

> *In circumstances R, is it the case that A expresses MV?*

For example, one could argue that carrying on a pregnancy is generally the responsible thing to do, unless in circumstances R where the pregnant person is underage (a direct attack on $Exemplifies Virtue(R, A, MV)$). From here, the inquiry can extend beyond the immediate situation to ask:

> *In circumstances R', would A still express MV?*

Here, R can also encompass broader societal or cultural conditions. This motivates a subsequent CQ addressing a well-known critique of Virtue Ethics: *cultural relativism* [**?**], which holds that different cultures

prioritise different virtues, thereby rendering judgments of actions as morally right or wrong relative to a specific *cultural* context. If we let R' represent an alternative cultural lens, the question becomes:

> *Would action A viewed through diverse cultural perspectives, still be interpreted as expressing MV?*

This reflects the fact that what might be seen as an expression of MV in one culture could be interpreted differently in another. For example, in some cultures, the action of addressing elders with formal titles is seen as exemplifying the virtue $MV = respect$. However, in other cultures, the same action may come across as overly distant or impersonal, and therefore would not be viewed as expressing the intended virtue.

The next CQ examines the *intention* of the agent performing action A. Virtue Ethics is fundamentally an *agent-based* theory [**?**], which evaluates moral actions not solely by their outcomes or adherence to rules, but by the character and motivation of the person acting. Accordingly, the moral worth of an action depends not only on its outward form but also on the virtuous disposition from which it arises. In contrast with consequentialists and deontologists that focus on agents' decisions, virtue ethicists also consider the agent's motives, reasoning, emotions, demeanour, and attitude to be relevant [**?**]. A significant critique, then, for a proposed action is that, while appearing virtuous, it is, in fact, performed to mainly signal virtue to others, commonly referred to as *virtue signalling* [**?**]. This leads to the following critical question:

> *Is action A motivated by a genuine disposition to express MV?*

Undermining the motive of the action essentially undermines the argument for this action in a Virtue Ethics framework. For example, consider again the argument advocating for donating to charity Z. A counter-argument might point out that the sole motive of the person suggesting action A is its public announcement on social media (ensuring widespread visibility), suggesting that the action was driven by reputational motives. This undermines the argument, attacking the premise f_4 about action A exemplifying MV.

The last CQ of this group examines whether there is an alternative action A' that exemplifies virtue MV to a greater degree; e.g., A_g: "A person is actually responsible when they recognise they cannot raise a child and choose not to". The critical question is formulated as,

> *In R, does performing action A preclude some other action A' that would express MV to a greater degree?*

If A indeed precludes some A' that exemplifies MV, the CQ results in the generation of two symmetrically attacking arguments A_1, A_2, instantiating the *Action Scheme from Virtue Ethics*:
A_1: $Circumstances(R)$, $ExemplifiesVirtue(R, A, VA, MV)$, $Effect(R, A, E)$, $Goal(R, A, G) \Rightarrow ShouldPerform(R, A)$
A_2: $Circumstances(R)$, $ExemplifiesVirtue(R, A', VA, MV)$, $Effect(R, A', E')$, $Goal(R, A', G')$ $\Rightarrow ShouldPerform(R, A')$[4]
This type of conflict should be resolved at the meta-level (see Section **??**) to determine which action better expresses the virtue.

Group 4: Other Virtues The fourth group of critical questions focuses on alternative virtues MV' and is constructed analogously to the third group. Firstly, I examine if action A (in circumstances R and with effect E) is in conflict with some other virtue MV'. One can think of examples where different interpretations of an action in terms of virtues can end up with conflicting views as to whether the action is right or not. Argument A_g is making the claim that abortion can be seen as expressing the virtue of responsibility. A conflicting interpretation though, is that this action is selfish. I introduce the following CQs to capture this:

[4]It can also be the case that the two arguments state the same goal (G'=G).

Is there a different interpretation of action A, according to which A is in conflict with some other virtue MV'?

and

Does performing action A have an effect E' that is in conflict with some other virtue MV'?.

One could then try to defend the original argument by coming up with alternative virtues the action expresses, answering the question:

Does performing action A, with effect E, also exemplify some other virtue MV'?

For instance, while an argument supporting A_g might acknowledge conflict of action $A = abortion$ with $MV = selflessness$, it could also advocate that the action expresses additional virtues, such as courage, alongside responsibility. This highlights the need for a scheme that addresses the concept of argument accrual (more on this in Section **??**).

Similarly to the previous set of CQs, I also include a CQ that examines alternative action(s) A'. This time, the CQ examines two mutually exclusive actions A, A' that express two different virtues MV, MV', respectively. This CQ illustrates a significant critique of Virtue Ethics: the *conflict problem*, i.e. the fact that different virtues might have requirements that conflict as they might point in opposing directions [**?**]. For instance, perhaps the virtue $MV = responsibility$ can be appealed for $A =$*Abolish Abortion*, however, the virtue $MV' = compassion$ might dictate supporting the individual's choice to have an abortion as a means of alleviating their suffering. So, I introduce the CQ:

Does performing action A preclude some other action A' that would exemplify some other virtue MV'?

These CQs lead to bidirectional rebuttals again, with A' and MV' yielding a counter symmetrical rebutting instance of the scheme. This conflict can be resolved by deciding which virtue is preferred, thus the follow-up discussion would focus on which virtues are more significant or more relevant for circumstances R, pointing to the need for prioritisation of virtues on the meta-level, discussed in Section **??**.

Group 5: Virtue MV Building on this last critical question, the fifth category of CQs focuses on the virtue *MV per se*, drawing primarily from two core philosophical critiques of Virtue Ethics: the epistemological problem that arises from the fact that it is an *agent-based* theory [**?**] and cultural relativism [**?**]. In light of cultural relativism, similarly to the CQs introduced in Group 3, I have developed two additional critical questions to examine the cross-cultural acceptance of virtues. The first question,

Is MV universally accepted as a virtue?

highlights the fact that virtues can vary significantly across different cultures. For example, filial piety, which is highly regarded in Japan, may not be as prominently valued in the Netherlands. This disparity in cultural values can significantly impact how pertinent an argument that appeals to $MV = FilialPiety$ is in different cultural settings. The second one is:

Is it the case that in circumstances R, MV is a virtue that needs to be exemplified?

Both these questions point to undercutters. Let us, for example consider argument U that attacks argument A_r by answering the latter CQ by using the premise f_2: *The virtue MV (responsibility) is irrelevant in the current context*. Argument U then undercuts A_r, since Conc$(U) = \neg\, d_v$.

What is more, to address the fact that Virtue Ethics is predominantly agent-based, rather than act-based, I incorporate CQs that focus on agents. The original *argument from action* is action based, hence no reference to agent; however, as previously explained, Virtue Ethics focuses on the type of person one should be and virtues are defined by references to such virtuous agents. As a result, this necessitates careful consideration when one applies Virtue Ethics to specific actions. Drawing from this, I construct a CQ that evaluates MV via said VA agents:

> *In circumstances R, are there examples of VAs considered to possess virtue MV who do not sanction/perform A?.*

This inquiry points to a potential undermining attack on premise f_4, seeking to verify the real-world alignment between VAs and action A.

Lastly, one significant criticism of Virtue Ethics states that virtue is neither necessary nor sufficient for right action [**?**]. I introduce the following CQ to illustrate this:

> *Is appealing to virtue MV a sufficient reason to perform action A in circumstances R?*

For example, consider an argument that claims: "A courageous person ($MV =$ Courage) should confront an armed intruder in their home to protect their family". While this action may indeed exemplify courage, appealing to courage alone may not be a sufficient reason to perform it, as there are risks involved. This points to an undercutting attack that explores whether the presence of a virtue alone justifies an action, examining if virtues by themselves are adequate to motivate the action.

Group 6: Other Ethical frameworks The point regarding sufficiency and necessity leads us to the final category of critical questions that focus on additional factors that might characterise an action as morally correct. This last category examines how well the action aligns with other ethical frameworks, starting with Deontology.

Let us consider the historical example of Rwanda, where a colonel faced the decision on whether to intervene and halt the atrocities, appealing to the moral virtue $MV = compassion$ as the $VAs = peacekeepers$ do. However, action $A = intervention$ was in direct conflict with the deontological norm of non-intervention mandated to UN forces. Thus, a pertinent question is:

> *Does action A violate some deontological obligation/prohibition DN in R?.*

Secondly, I evaluate whether the proposed action aligns with consequentialist principles. In the case of the Rwandan Genocide, the intended outcome of action $A = intervention$ was $E = GenocideHalt$. Group 2 addressed whether a proposed action would achieve its intended outcome (i.e., whether the enactment of a compassion-driven action will indeed result in the cessation of atrocities). Here, I take a step further by critically evaluating both the efficacy of the action dictated by the virtue $MV = compassion$ in general and whether, considering all possible consequences, this was truly the optimal course of action in the following way.

Firstly, I examine the action's actual outcomes:

> *Considering all possible consequences, is the proposed action A indeed the best course of action in circumstances R?*

Secondly, I assess the motivation behind the virtue-driven action altogether:

> *Does acting in ways that express the virtue MV lead, on balance, to good consequences?*

Consider the argument: "One should remain silent about their company's misconduct because loyalty is a virtue, and whistleblowing betrays the team". Here, the proposed action $A =$ "stay silent" is said to exemplify the virtue $MV =$ loyalty. However, a counter-argument could say: "Routinely acting on loyalty has enabled harmful behaviour to persist unchecked throughout history".

Lastly, one could examine the compatibility of the proposed action with Rights-based Ethics. This approach involves questioning the alignment of the action of intervention with internationally recognised human rights T. A CQ in this context is:

Is there a conflict between action A and a universally established right T in R?

In the Rwanda example, such a question assesses whether intervening to stop the genocide conflicts with, or supports, the human rights established by international law (e.g., the right to life, right to protection against genocide, etc).

Summary Below, I summarise all CQs for an *action scheme from Virtue Ethics*.

1. Circumstances & Feasibility
   - Are circumstances R true?
   - Are circumstances R plausible/possible?
   - Are circumstances R as stated complete?
   - Is action A possible in circumstances R?
2. Effects E & Goal G
   - Does action A have effect E in circumstances R?
   - Will action A realise goal G in circumstances R?
   - In R, does performing action A have a side effect E' that is in conflict with the moral virtue MV?
3. Link between A & MV?
   - (How) does action A exemplify the virtue(s) MV identified as relevant in circumstances R?
   - In circumstances R, is it the case that A expresses MV?
   - In circumstances R', would A still express MV?
     - Would action A viewed through diverse cultural perspectives, still be interpreted as expressing MV?
   - Is action A motivated by a genuine disposition to express MV?
   - In R, does performing action A preclude some other action A' that would express MV to a greater degree?
4. Other Virtues MV'
   - In R, is there a different interpretation of action A, according to which A is in conflict with some other virtue MV'?
   - In R, does performing action A have a side effect E' that is in conflict with some other virtue MV'?

- In R, does performing action A, with effect E, also express some other virtue MV'?
- In R, does performing action A preclude some other action A' that would express some other virtue MV'?

5. Virtue MV
   - Is MV universally accepted as a virtue?
   - Is appealing to virtue MV a sufficient reason to perform action A in circumstances R?
   - Is it the case that in circumstances R, MV is a virtue that needs to be exemplified?
   - In circumstances R, are there examples of VAs considered to possess virtue MV who do not sanction/perform A?
6. Other Ethical Frameworks
   - Does action A violate some deontological obligation/prohibition DN in R?
   - Considering all possible consequences, is the proposed action A indeed the best course of action in circumstances R?
   - Does acting in ways that express the virtue MV lead, on balance, to good consequences?
   - Is there a conflict between action A and a universally established right T in R?

4 Action Scheme from Deontology

In ethical discourse, notions like principles and rules occupy a central role; in the abortion debate, for instance, religious norms are frequently cited. Consider the following argument from the *Kialo* debate *Pro-Life vs Pro-Choice: Should Abortion be Legal?*.

> A_c: "Contraception is considered immoral by certain religions because it interferes with the conception of a child. Religions who are against contraception are therefore also against abortion because it interferes with the birth of a child."

To capture arguments that cite norms and rules, I develop this scheme, modelling arguments that appeal to Deontology. For this, I adapt *argument from action* as follows.

- I cite deontological norms DN.
- I distinguish between obligation, permission, and prohibition and create three sub-schemes.
- I set the motivation for an action to be directly linked to a norm DN.
- I do not include a goal G as a variable. I assume the goal to be constant $\mathcal{G}$ = "adherence to DN", since in deontological reasoning the justification for an action lies in its conformity to the norm DN (and not in the achievement of a particular outcome).

I construct the *Action Scheme from Deontology* as follows. Similarly to constructing the *Action Scheme from Virtue Ethics*, the first thing the scheme takes into account is the current circumstances R that necessitate the consideration of the respective deontological norm. The following step is identifying the specific deontological norm that is applicable in the current situation: this norm could be an obligation (something

we must do), a prohibition (something we must not do), or a permission (something we are allowed to do). Thus, I introduce three variations of *Action Scheme from Deontology*, along with their corresponding critical questions. In formulating said schemes and their CQs, I use the following terminology.

- Action A *adheres to/comforms with/obeys* deontological norm DN if it complies with it.
- Otherwise, action A *violates/disobeys* deontological norm DN.

4.1 Prohibition

Box 1 Action Scheme from Deontology (prohibition)

In the current circumstances R
Deontological norm DN_p is applicable
Therefore, we should **not** perform action A
{Which has an effect E}
Because action A is prohibited under DN_p

First, I introduce the sub-scheme for prohibition in Box **??**, encapsulating deontological arguments citing prohibitive norms, such as "We should ban abortions because one should not kill". The bracketed line is optional: the effect E is not the motivational basis for refraining from action A; however, it is *causally linked* to A and thus becomes relevant in the broader argumentative context/dialogue. Even if the original proponent of the argument does not explicitly mention E, interlocutors may introduce considerations about it. Since critical questions related to E should be included to properly evaluate an argument that falls under this scheme, I include a variable corresponding to effects in the formulation of the scheme to make it available for structured analysis.

The aforementioned example of a deontological argument is an argument sourced by *Kialo*, again, so it is a user-authored argument positing that abortions should not be performed. A refined example to illustrate the deontological scheme expressing prohibition more formally is the following, where $R = Pregnancy$, $DN_p = ShouldNotKill$, $A = Abortion$, $E = EndOfFetus$.

> A_p: "In the case of a pregnancy, the normative rule 'one should not kill a human being' applies. One should, therefore, not perform an abortion (which kills the fetus) because that is forbidden under the aforementioned norm."

A representation of the scheme's premises in *ASPIC*$^+$ is then:
f_1 : *We are in circumstances R.*
f_2 : *In R, the prohibitive deontological norm DN_p is applicable.*
f_3 : *Action A is prohibited under the deontological norm DN_p.*
$\{f_4$: *In R, A has effect E.*$\}$

The defeasible rule of the scheme can be represented as follows.
d_p (R, DN_p, A, E): $Circumstances(R)$, $Applicable$ (R, DN_p), $Prohibit$ (DN_p, A), $\{Effect(R, A, E)\}$ $\Rightarrow ShouldNotPerform(R, A)$

So, argument A_p becomes:

$$A_p : [f_1, f_2, f_3, \{f_4\}] \Rightarrow_{d_p} ShouldNotPerform(R, A)$$

4.1.1 Critical Questions

Similarly to how I proceeded for the *Action Scheme from Virtue Ethics*, I now introduce a comprehensive set of critical questions for the first sub-scheme of the *action scheme from Deontology*.

Group 1: Circumstances and Feasibility The first set of questions is the same as in the *Action Scheme from Virtue Ethics*. I incorporate the first group of CQs as described in Section **??**.

Group 2: Relevance and Link I introduce additional critical questions that assess the relevance and link of the cited deontological norm DN_p to circumstances R and action A. The first question challenges $Applicable$ (R, DN_p) and the jurisdictional scope of the norm DN_p, focusing on whether appealing to DN_p is relevant and whether DN_p governs the domain where it is invoked. Such issues can arise, for example, when religious norms of a specific religion are applied to individuals who do not belong to that religion (e.g., an argument advocating abstraining from pork based on the religious norm $DN_p = DoNotConsumePork$). In the running abortion example, this would correspond to an argument against abortion based on a religious norm DN_p that forbits it. A counter-argument would contest whether such a norm is binding on those outside the religious community. A different example can be found if I consider the different legislations accross countries. For instance, consider $R =$ "Living in France" and a norm from UK law DN_p ="One must not drive on the left side". For an argument from deontology for action $A =$ "Driving on the right" based on this DN_p, a counter-argument would state that the norm has no authority in this context. Even at a simpler level, this is illustrated by everyday "house rules"; a rule such as $DN_p =$ "No shoes indoors" may be binding within one household and not carry any normative force in another. Such arguments attack premise f_2, challenging whether the stated principle is relevant or applicable. I introduce the following CQ to challenge the relevance of DN_p in the specific context R.

Is the deontological norm DN_p applicable in the circumstances R?.

Building on this, by extension of the applicability of the cited norm, I introduce a question that examines if the prohibited action is, in fact, prohibited by DN_p. Even if one accepts that the norm is applicable, this second question explicitly probes whether the particular action under discussion violates the norm. As shown in the aforementioned examples, the abortion debate often revolves around arguments that refer to the principle $DN_p =$"One should not kill". However, this argument is contested by claims advocating that terminating a fetus at an early stage does not qualify as killing, since a fetus is not (yet) alive. Accordingly, I formulate the following CQ, which points to undermining attacks on premise f_3.

Does action A (unambiguously) violate norm DN_p in R?

The two questions can overlap in some contexts depending on argument framing. For example, one might argue in the abortion debate that $DN_p =$ "One should not kill" applies only to born human life; in this case, the objection is, again, about applicability.[5] However, this second question mainly targets definitional boundaries; i.e., does this action fall within what the norm prohibits? To illustrate the distinction further, consider the earlier case of driving in France. Driving on the right "technically violates" $DN_p =$'Do not drive on the right, but DN_p is inapplicable. By contrast, consider a case where someone locates their stolen bicycle and chooses to reclaim it. One could claim that action $A = RetrieveBike$ is prohibited by an appeal

[5]Note that 'applicability' here refers to whether the circumstances fall within the norm's conceptual boundaries, not territorial jurisdiction. Arguing that 'do not kill' is inapplicable in these circumstances treats personhood as a boundary *condition* that defines the norm's scope, rather than challenging the norm's authority *per se*.

to the norm DN_p = "One should not steal". A counter-argument to this would point that this action does not, in fact, violate the cited norm as stealing is, by definition, the unauthorised taking of another's property (but it is an applicable norm). This distinction mirrors legal reasoning, where courts must first determine if legal provision or precedent applies to the type of case at hand (jurisdictional scope), and then whether the facts satisfy the provision's/precedent's specific criteria (definitional boundaries). Moreover, arguments that invoke exceptions to DN_p can function as responses to this CQ; although the rule remains applicable, the action in question does not constitute a violation because it falls under a recognised exception. For example, consider the regulatory norm DN_p = "Vehicles are prohibited from parking in this designated area" with an explicit exception clause permitting parking on Sundays; in $R = Sunday$, parking in said area does not violate the norm. More generally, exceptions to rules play a significant role in normative reasoning. References to such exceptions are also found in the next group of CQs, as well as in Sections **??** and **??**.

Moreover, following a similar strategy to the one used in the Virtue Ethics framework, I introduce a critical question that examines whether refraining from action A leads to side effects that themselves violate the cited deontological norm DN_p.

Does refraining from action A, in circumstances R, have a side effect E' that violates DN_p?

This is a question that points to undercutting attacks. For instance, in arguments against prohibiting abortion in life-threatening situations, one might claim that preventing the woman from accessing an abortion leads to E' = her death; an outcome that itself can be considered as a violation of the norm DN_p = "One should not kill". So, in circumstances $R = LifeThreateningPregnancy$, refraining from action $A = Abortion$ results in effect $E' = DeathofWoman$, which contradicts norm DN_p. This functions as an undercutter—it challenges the applicability of the norm to prohibit action A in the given circumstances R, on the grounds that such prohibition leads to a violation of the cited norm. To represent this formally, I model the circumstsances *life-threatening pregnancy* as shown in Section **??**, setting R = *Pregnancy*, $R' = LifeThreateningPregnancy$, and $C(R') \equiv \big(C(R) \wedge LT(R, W)\big)$ (where $LT(R, W)$ is a predicate for life-threatening pregnancy for the woman W). The prohibition of the action in circumstances R' leads to the death of the pregnant woman: $\neg\, A \wedge R' \rightarrow E'$. This results in the prohibition of the action violating norm DN_p that was the motivation for prohibiting the action, thus undercutting the argument supporting the prohibition. Formally: $C(R)$, $LT(R, W)$, $\big(C(R') \rightarrow \text{Violates}(DN_p)\big) \Rightarrow \neg\, d_p\ (R,\, DN_p,\, A,\, E)$

Group 3: Deontological Norm DN_p The next group of critical questions assesses the legitimacy of the cited norm *per se*. The first question engages the standard critique of Deontology that arises from *cultural/religious specificity* and *relativism* [**?**]: deontological reasoning is often challenged on the grounds that its supposedly universal norms may not reflect cultural or situational nuances and may not enjoy broad legitimacy outside particular religious or ideological communities. Norms presented as universal may in fact lack force outside the community or tradition from which they originate. For instance, arguments challenging A_c by questioning the use of religious beliefs as decisive factors in secular legal or political contexts exemplify this critique. A CQ reflecting this is the following.

How universally accepted is DN_p within the relevant ethical, moral, or legal framework?[6]

This question builds on the earlier notion of norm applicability and extends it to investigate its universality. Applicability concerns whether a norm formally governs the domain at hand (its technical or jurisdictional scope). By contrast, this CQ probes the *universality* or legitimacy of the norm, i.e. whether it *should* govern in the given context and whether its standing is recognised broadly rather than being confined to a

[6]The 'relevant ethical, moral, or legal framework' could, again, be viewed as generalised R.

specific cultural or ideological framework (mirroring the distinction of natural and positive law discussed in Section **??**). This question points to undercutting attacks that challenge the scheme's inference rule d_p. Returning to the example regarding religious norms in secular contexts, the distinction can be illustrated through different counter-arguments; A_{ap} : "This religious norm does not apply here because this is in a secular legal domain" (challenging the applicability of DN_p) vs A_{un}: "This religious norm should not be used in public policy because it lacks broad acceptance in our pluralistic society" (challenging the universality of DN_p). The sharpness of this distinction depends partly on the type of norm being invoked. Legal norms typically have well-defined jurisdictional boundaries, making applicability challenges more straightforward, whereas moral and religious norms often have contested scope and authority.[7]

Subsequently, this relates to a well-known philosophical challenge of deontological theories, regarding the distinction between *causing* harm and *failing to prevent* it [**?**]. Deontology typically distinguishes between actively bringing about harm through one's actions, i.e. doing harm, and letting harm happen by not intervening or not acting to prevent it, i.e. allowing harm. The critique that arises is that this distinction might seem arbitrary or morally problematic, since Deontology often judges actively causing harm as morally worse than allowing harm to occur, even if the outcomes (in terms of overall harm) might be identical or similar. The moral permissibility of abortion under certain deontological frameworks could depend heavily on whether abortion is classified as actively ending a life (killing), which is morally problematic as it violates DN_p ="One should not kill", or as failing to bring a life into existence (omission), which is potentially permissible, since failing to create new life is not equivalent to causing harm. Thus, if abortion is interpreted not as causing the end of a life but as *failing to bring a life into existence*, then the action may be seen as permissible within certain deontological frameworks. This idea is captured in the following CQ.

Could action A be re-framed such that it no longer violates norm DN_p?

Lastly, a common point that arises is whether norm DN_p is being applied consistently across analogous actions, or if its application is selectively targeting action A. For example, a common counter-argument to "abortion should be prohibited on the grounds that it prevents life" is to point out the inconsistency with other accepted practices, such as contraception or abstinence. One such argument from the ***Kialo*** debate proclaimed, "By that logic, contraception should be made illegal as well". As noted earlier, one of the roles of critical questions is to elicit contextual details that, in a dialogical setting following the Socratic method, could expose inconsistencies. For example, arguments that invoke biblical norms are frequently challenged by citing other biblical passages to demonstrate selective application. Take, for instance, an argument against homosexuality citing the norm $DN_p =$ "You shall not lie with a male as with a woman" from Leviticus [**?**] countered by a different argument referencing other Levitical prohibitions that are widely disregarded today (e.g., $DN_p =$ "[...] neither shall a garment mingled of linen and woollen come upon thee" [**?**]). I introduce the following CQ to point to such challenges.

Are there exceptions to the deontological norm DN_p that are accepted in circumstances R?.

This type of question leads to challenges about the consistency of the normative framework being invoked, suggesting that if one biblical prohibition is treated as binding, logical consistency requires either accepting all such prohibitions or providing principled reasons for selective application. This shifts the discourse to a meta-level discussion and the conflicts of these arguments require resolution through meta reasoning.

[7]This maps onto the philosophical distinction between *duties of right* and *duties of virtue* (see Table **??**), where the former are enforceable and have clearer external boundaries, while the latter depend more on internal moral frameworks.

Group 4: Other Norms DN' The next set of critical questions addresses potential conflicts of action A with other deontological norms. Said questions draw from one of the most prominent criticisms of Deontology, and a broader challenge for all ethical theories: moral dilemmas. Of course, deontological norms often conflict with one another. For instance, consider the norm $DN_p =$ 'One should not harm another' and another norm $DN_o =$ 'One should protect oneself'. In circumstances $R =$ 'I am being threatened', these two norms point toward conflicting actions. Such dilemmas arise frequently in everyday reasoning and multiple examples can be found in applied ethics (e.g., in biomedical ethics, in business ethics, or in legal ethics). To reflect this, I introduce the following question.

> *In circumstances R, is adhering to DN_p, by refraining from action A, in conflict with some other deontological norm DN'?*

This question can be further refined into the following three sub-questions. First, two questions that address the conflict of a prohibitory norm with an obligation. To address the aforementioned scenario, I introduce:

> 1) *In circumstances R, is action A obliged by some other deontological norm DN_o?*

The second question examines whether refraining from an action *precludes* performing a different action that is itself obligatory under another norm. Let us consider the norms $DN_p =$"One should not lie" and obligation $DN_o =$ "One must protect the innocent" (see Section **??** for the scheme appealing to obligations) in the classic 'murder at the door' scenario [**?**], where $R =$ "A murderer asks where an intended innocent victim is hiding". Here, DN_p forbids action $A =$ "Lie to the murderer", while DN_o obliges $A' =$ "Protect the potential victim", two actions that in the current circumstances appear incompatible, thus I introduce the question:

> 2) *Does refraining from action A preclude some other action A' obliged under some other deontological norm DN_o?*

In both these cases, the CQs give rise to arguments instantiating *Action Scheme from Deontology* that rebut each other.

The next sub-question targets harder cases of conflicts that arise when two negative duties are in conflict; cases that abstaining from an action as demanded by a prohibitory norm DN_p effectively forces the agent to perform a second action that is itself forbidden by another norm DN'_p. Consider, for example, a corporate lawyer who is served with a subpoena. The lawyer faces two conflicting negative duties: a duty of confidentiality $DN_p =$ "Do not disclose a client's secrets" and a duty to the court $DN'_p =$ "Do not refuse a lawful subpoena". In this scenario, avoiding one forbidden action effectively necessitates committing another. I introduce the following critical question.

> 3) *Does refraining from action A imply some other action A' prohibited under some other deontological norm DN'_p?*

Addressing all the aforementioned conflicts necessitates prioritisation of norms on a meta level, which is addressed in Section **??**.

Lastly, analogously to the last question of the third group, I examine whether a reformulation of action A could be in fact permitted or obliged under a different norm DN'.

> *Could action A be re-framed so that it is permissible/obliged under a different norm DN'?*

Group 5: Other Ethical Frameworks In the last group of CQs, I introduce the CQs that examine the compatibility of the argument with the other ethical frameworks. Each of these questions invites counter-arguments that rebut arguments that appeal to Deontology by drawing on premises from alternative ethical frameworks.

Firstly, I examine whether refraining from action A is in conflict with a moral virtue. Consider a scenario where strict adherence to a religious norm DN might be in conflict with the moral virtue of empathy. For example, on a Sabbath afternoon, a neighbor collapses in the street in clear need of urgent assistance. Based on the religious norm DN_p that prescribes that all forms of work are prohibited on Sabbath, an *Action Scheme from Deontology* advocates refraining from action $A = Aid$. A counter-argument, then, appealing to Virtue Ethics could argue for action A based on the virtue of empathy. A CQ addressing this is:

In circumstances R, is prohibiting action A in conflict with moral virtue MV?

Secondly, I investigate the effects E of action A and evaluate the implications of refraining from it through a consequentialist lens. One of the most prominent criticisms of Deontology is its rigidity and its failure to account for the (full scope of) consequences [**?**]. Blind adherence to normative rules can sometimes lead to disastrous outcomes. A frequently cited philosophical dilemma illustrates this tension: the "ticking time bomb" scenario [**?**], where strict adherence to the deontological norm $DN_p =$ "One should not torture" might result in (preventable) deaths of thousands. To address such concerns, I firstly pose a critical question that considers the broader utility of the norm in general (one that also relates to the first question in the third group of CQs, which inquires about the universal acceptability of DN_p).

Does following norm DN_p lead on balance to positive consequences?

Then, I investigate the action's effects E in R. Since they are not necessarily explicitly stated in the scheme, the following CQ prompts to articulate them.

What are the effects E of refraining from action A in circumstances R?

and subsequently

Is E an on balance positive consequence in circumstances R?

These questions focus on the consequences of applying the norm DN_p in the specific context R. For instance, in the scenario of the captured terrorist, an argument against torture will appeal to the prohibition $DN_p =$ "One should not torture". A counter-argument could then be: "Although in general it is right not to torture, in this case thousands of lives are at immediate risk. Therefore, the positive outcomes of extracting information outweigh the moral cost of disobeying the norm".

Lastly, analogously to the frequent phenomenon of conflicting duties, conflicts arise between duties and rights as well. The following question encapsulates this.

In circumstances R, is refraining from action A in conflict with universally established right T?

Summary Below, I summarise the critical questions for an *Action Scheme from Deontology* that focuses on prohibition.

1. Circumstances & Feasibility
    - Are circumstances R true?
    - Are circumstances R plausible/possible?

- Are circumstances R as stated complete?
- In circumstances R, is refraining from action A possible?

2. Relevance & Link
    - Is the deontological norm DN_p applicable to circumstances R?
    - Does action A (unambiguously) violate norm DN_p in R?
    - Does refraining from action A in circumstances R have a side effect E' (dictates A') that itself violates norm DN_p?
3. Norm DN_p
    - How universally accepted is DN_p within the relevant ethical, moral, or legal framework?
    - Could action A be re-framed such that it no longer violates norm DN_p?
    - Are there exceptions to the deontological norm DN_p that are accepted in the current circumstances R?
4. Other Norms DN'
    - In circumstances R, is adhering to DN_p, by refraining from action A, in conflict with some other deontological norm DN'?
        - In circumstances R, is action A obliged by some deontological norm DN_o?
        - Does refraining from action A preclude some other action A' obliged under some deontological norm DN_o?
        - Does refraining from action A dictate some other action A' prohibited under some other deontological norm DN'_p?
    - Could action A be re-framed so that it is permissible/obliged under a different norm DN'?
5. Other Ethical Frameworks
    - In circumstances R, is prohibiting action A in conflict with moral virtue MV?
    - Does following norm DN_p lead on balance to positive consequences?
    - What are the effects E of refraining from action A in circumstances R?
        - Is E an on balance positive consequence in circumstances R?
    - In circumstances R, is refraining from action A in conflict with universally established right T?

4.2 Obligation

Similarly, I create an argument scheme that captures arguments grounded in obligations, such as "We should legalise abortions because one is obliged to respect bodily autonomy". This type of reasoning frames the right to abortion as grounded in a moral obligation to uphold individual autonomy and personal agency.[8] As previously discussed, Deontology distinguishes between *negative duties*, which prohibit actions, and *positive duties*, which require actions. Negative duties, such as the prohibition against harming others, are

[8] It is worth noting that this argument can also be interpreted as an *Action Scheme Appealing to Rights* (see Section **??**), where the focus is specifically on the protection of fundamental, historically earned entitlements.

typically clear-cut and more often align with perfect duties: they demand strict compliance and allow little latitude. Positive duties, in contrast, can sometimes be perfect (e.g., the obligation to "appear in court when subpoenaed") but frequently manifest as wide, imperfect duties, such as "one should care for fellow humans" (see Table **??**), leaving substantial room for interpretation regarding the precise actions they require. Because of this, negative duties are often considered stronger, more immediate, and morally clearer than positive duties. For instance, while one might have a moral obligation to actively help others (a positive duty), the obligation to refrain from causing harm (a negative duty) is usually regarded as more urgent, direct, and binding. Positive duties require one to actively pursue specific ends, and thus their associated actions are often subject to greater debate about their exact scope (and urgency).

This distinction mirrors a key tension in consequentialist ethics between *maximising* and *satisficing* approaches (see Section **??**). Satisficing consequentialism holds that an action is morally acceptable so long as it produces a sufficient amount of good, meeting a threshold is enough. By contrast, maximising consequentialism demands that one always pursue the action that produces the greatest possible overall good. Although this parallel does not have exact equivalences, and one could argue the analogy works in either direction depending on the interpretative lens,[9] rather than a direct mapping, it strives to demonstrate that the two ethical frameworks are confronting similar dilemmas regarding moral scope, action guidance/specificity and perfection/demandingness. Within this structure, the concept of *supererogation* [**?**] arises: actions that exceed what is strictly required by an imperfect duty are seen as morally praiseworthy but not obligatory. However, this raises difficult questions about where exactly the threshold for 'enough' lies, leaving considerable room for interpretation and debate.

With these considerations in mind, I develop the *Action Scheme from Deontology (obligation)*, presented in Box **??**.

Box 2 Action Scheme from Deontology (obligation)

In the current circumstances R,
Deontological norm DN_o is applicable
Therefore, we should perform action A
{Which has an effect E}
Because action A is obliged under DN_o

A representation of the scheme's premises in *ASPIC*$^+$ is then:
f_1 : *We are in circumstances R.*
f_2 : *In R, the obligatory deontological norm DN_o is applicable.*
f_3 : *Action A is dictated under the deontological norm DN_o.*
$\{f_4$: *In R, action A has effect E.*$\}$

Additionally, the defeasible rule can be represented as follows.
d_o (R, DN_o, A, E): $Circumstances(R)$, $ApplicableRules$ (R, DN_o), $Dictates$ (DN_o, A) $\{Effect(R, A, E)\} \Rightarrow ShouldPerform(R, A)$

[9] From one perspective, negative, perfect duties resemble satisficing in that fulfilling a minimal condition is enough (e.g., not killing), whereas imperfect (often positive) duties mirror maximising, encouraging continuous moral effort (e.g., promoting the good). From another viewpoint, perfect duties are more like maximising since partial compliance is unacceptable (e.g., "mostly not killing" is meaningless), while imperfect duties are more like satisficing, where any contribution is morally better than none, but failure is not always blameworthy.

So, argument A_o becomes:

$$A_o : [f_1, f_2, f_3, \{f_4\}] \Rightarrow_{d_o} ShouldPerform(R, A)$$

4.2.1 Critical Questions

Below, I explicitly highlight the aspects that remain consistent with the *prohibition* scheme, as well as identify areas where adjustments or additions are necessary to appropriately address the distinct nuances associated with positive obligations.

Group 1: Circumstances and Feasibility For this group, I retain the critical questions as formulated in the prohibition scheme. Notably, the fourth question, which concerns the feasibility of an action, takes on greater importance here, since performing an action may be more difficult than refraining from one.

Group 2: Relevance and Link For obligations, the critical question from the second group assessing whether action A adheres to the cited norm DN becomes even more central. This is because, unlike prohibitions, obligations often involve greater interpretive flexibility, making the justificatory link between norm and action less direct. To enhance scrutiny in such cases, I introduce an additional question:

> *Are there alternative actions A that fulfil the obligation under DN_o more effectively or appropriately?*

For example, imagine argument A_{rel}, advocating for action $A = RelocateAbroad$, based on the obligation $DN_o = MinimiseCarbonFootprint$

f_1 :	*Factories emit large amounts of CO_2*	(R)
f_2 :	*Firms must minimise their carbon footprint*	(Applicable(R, DN_o))
f_3 :	*Relocating production abroad is obligated by this rule*	(Dictates(DN_o, A))
f_4 :	*Domestic emissions drop to near-zero*	(Effect(R, A, E))
A_{rel} :	$[f_1, f_2, f_3, f_4] \Rightarrow$ *ShouldPerform*(R, A)	

A counter-argument responding to the aforementioned CQ, and advocating for an alternative action $A' = OnsiteInstallations$, could involve a different f_3 premise, f_3': "Installing on-site solar is dictated by DN_o", along with a different corresponding $Effect(R, A', E')$ (i.e., f_4'). This argument would then lead to the conclusion $\therefore$ $ShouldPerform(R, A')$, rebutting A_{rel} by attacking its conclusion. What is particularly noteworthy here is that the optional placement of the effect variable may become relevant. Even within a deontological framework, the rebutting argument might assert that the alternative action A' fulfils the norm DN_o more effectively (or with *more positive* or *fewer negative* side-effects) than the originally proposed action A. This type of conflict once again points to the necessity of resolution at the meta-level, where competing instantiations of the same norm must be comparatively assessed, often by resorting to consequentialist reasoning.

Group 3: Deontological Norm DN_o I maintain the CQs from the prohibition scheme, and I additionally introduce new CQs inquiring about a specific type of obligations, contrary-to-duty (CTD) obligations. CTD obligations are conditional: they only become relevant when a primary norm DN' has been violated. I

introduce a CQ for when the cited DN_o is being invoked as a CTD obligation, in order to test whether DN_o is truly CDT (i.e., whether the triggering condition of prior norm violation has actually occurred, or whether it is merely being assumed or strategically exaggerated).

> *Is DN_o a CTD obligation in R?*

The purpose of this question is to prevent premature invocation of CTD obligations as a means to effectively violate the primary norm DN'. The CQ points to an undermining attack, challenging the applicability of the invoked norm (premise f_2). For example, an argument regarding climate change, advocating for action A = "focus on adaptation measures" and invokes the contrary-to-duty obligation DN_o = "When mitigation fails, adaptation measures must be implemented". This CQ examines whether the mitigation targets truly failed, or whether this is a premature concession.[10]

Building on this, I target a distinctive subclass of CTD obligations: *corrective duties* [**?**]. These obligations function as a direct response to a prior violation of a different norm; they require an agent to repair, compensate, rectify the effects of said violation. For example, in "Companies responsible for pollution must fund the restoration of affected ecosystems", the obligation is not grounded in a wide duty to protect the environment, but in a narrow duty to *rectify harm already caused*. I capture this line of reasoning with the following question:

> *Is the cited obligation under DN_o a corrective duty arising from a prior violation of a different norm DN'?*

This question is important for distinguishing between forward-looking obligations and reparative ones, which may carry different justificatory weight or urgency. Clarifying whether an obligation is corrective in nature may resolve ambiguity about its normative strength. An action that initially appears to reflect an *imperfect*, supererogatory duty (e.g., "This company should impose green policy X because we should care about the environment") might invite objections such as: "Policy X is too strict, the company is already caring substantially about the environment". One might then strengthen the original argument by questioning the nature of the obligation, resulting in an argument appealing to a *perfect*, reparative obligation that the agent is morally bound to fulfil (e.g., "This company should adopt green policy X as a necessary response to the environmental damage it has caused").[11] Additionally, this CQ also helps surface contextual information that may become relevant at the meta-level of analysis (see Section **??**), where competing duties are weighed against one another.

Group 4: Other Norms DN'_o I slightly reformulate the CQs for prohibition to fit within a context of obligation, beginning with the general guiding question, followed by its subsequent sub-questions.

> In circumstances R, is performing action A to adhere to DN_o violating some other deontological norm DN'?

[10] This CQ can also be seen as a specific instance of the first group of CQs, which inquire the truthfulness and completeness of circumstances R; in this case, R = "Climate mitigation targets have been missed". However, I treat it as a distinct case, since the CQ specifically targets the violation of a prior norm, not just factual conditions. This inquiry is of great significance for decision-support systems, which might otherwise shift too quickly into fallback or non-ideal modes without sufficient justification. Additionally, invoking a CTD obligation can sometimes be used to minimise responsibility while ignoring the original harm.

[11] In many legal systems, a corrective requirement is formally expressed as a new, separate norm; for example, DN'_o = "Whoever pollutes must fund remediation" might be considered different to the more basic DN_o ="Do not pollute". Conceptually, however, DN'_o is *derivative*: its normative force depends on DN_o. This distinction highlights a broader modelling consideration: whether to codify wide duties and their more narrow operationalisations as distinct norms. The appropriate approach may depend on the needs of the application domain. Given that the authority of DN'_o is parasitic on the breach of DN_o, and following the grouping adopted here, I classify this critical question under Group 3, which focuses on the nature, status, and origin of the cited deontological norm.

This gives rise to the following subquestions.

1) *In circumstances R, is action A prohibited by some norm DN_p?*

For example, a catering employer might wish to donate leftover buffet food to the homeless ($A = Donate$) out of a religious obligation DN_o to feed the poor. However, this action could be prohibited by a policy norm DN_p that reflects the catering company's own internal policy.

2) *In circumstances R, does performing action A dictate some other action A' prohibited by some DN_p?*

For example, an IT manager might wish to donate a batch of retired company laptops to a local school ($A = Donate$) in order to reduce e-waste and support digital inclusion. However, the company's data-security policy stipulates that any device leaving the premises must first be wiped to a certain standard. Hence performing A dictates a further action $A' =$ "carrying out the level X wipe", $DN_o(A) \rightarrow DN_o(A')$. Yet, a policy DN_p prohibits staff from conducting such high-level wipes without written approval from the Chief Information Security Officer. Thus, in these circumstances, A' is prohibited by DN_p.

3) *In circumstances R, does performing action A preclude some other action A' obliged by some DN'_o?*

For example, during the COVID-19 pandemic a public-health order DN_o obliges people who have tested positive to self-isolate at home for ten days ($A = SelfIsolation$). At the same time, a single parent has a standing legal duty of care DN'_o to collect and supervise their young child after kindergarten each afternoon ($A' = CollectChild$). Performing A (strict self-isolation) precludes performing A' (collecting the child), illustrating the conflict this CQ points at. Addressing all the aforementioned conflicts necessitates prioritisation of norms on a meta level, which is, again, addressed in Section **??**.

Group 5: Other Ethical Frameworks For the final group, the critical questions remain largely the same, with only minor adjustments to reflect the shift from prohibitive to obligatory reasoning.

Summary Below, I summarise all the critical questions for the *Action Scheme from Deontology* that focuses on obligations.

1. Circumstances & Feasibility
    - Are circumstances R true?
    - Are circumstances R possible?
    - Are circumstances R as stated complete?
    - Is action A possible in circumstances R?
2. Relevance & Link
    - Is the deontological norm DN_o applicable to the current circumstances R?
    - Does action A adhere to DN_o (unambiguously) in R?
    - Does performing action A in R have a side effect E' (dictates an A') that violates DN_o?

- Are there alternative actions A that fulfil the obligation under DN_o more effectively or appropriately, in R?

3. Norm DN_o
   - Is DN_o a CTD obligation in R?
   - Is the cited obligation under DN_o a corrective duty arising from a prior violation of a norm DN'?
   - How universally accepted is DN_o within the relevant ethical, moral, or legal framework?
   - Could action A be re-framed such that it is no longer obliged under DN_o?
   - Are there exceptions to the deontological norm DN_o that are prohibited in the current circumstances R?
4. Conflicts with DN'
   - In circumstances R, does performing action A, to adhere to DN_o, violate some other deontological norm DN'?
     - In circumstances R, is action A prohibited by some norm DN_p?
     - In circumstances R, does performing action A entail some other action A', prohibited by some DN_p?
     - In circumstances R, does performing action A preclude some other action A', obliged by some other DN'_o?
   - Could action A be re-framed so that it is prohibited under some norm DN_p?
5. Other Ethical Frameworks
   - Is performing action A in conflict with moral virtue MV?
   - Does following norm DN_o lead on balance to positive consequences?
   - What are the effects E of action A in circumstances R?
     - Is E an on balance positive consequence, in circumstances R?
   - In circumstances R, is performing action A in conflict with some universally established right T?

4.3 Permission

Lastly, I introduce a deontological scheme for *permission*, which encompasses arguments centred on morally permissible actions. In Deontology, permission typically indicates that an action falls within the boundaries of what is ethically acceptable, though it does not carry the normative force of obligation. That is, one *may* perform the action without committing a moral wrong, but is not required to do so. In the context of the abortion debate, for instance, arguments often acknowledge a general prohibition (e.g., against killing) but carve out exceptions based on competing deontological considerations. Consider the following example:

> A_{exc}: "Abortions should be permissible in cases of risk to the mother's health because one is allowed to prioritise one's own life."

This type of argument typically arises in response to critical questions about exceptions to general rules (e.g., *Are there exceptions to the deontological norm DN that are accepted in the current circumstances R?*). It helps differentiate between *obligatory* and *merely permissible* actions, acknowledging that certain acts, though not mandated, are ethically allowed under specific conditions.

Additionally, permission-based reasoning can bring internal tensions within deontological frameworks to light. In A_{exc}, the permissible act of abortion conflicts with a general prohibition, and thus raises issues of normative priority and exception-handling, questions which ultimately point to the need for meta-level resolution.

Box 3 Action Scheme from Deontology (permission)

In the current circumstances R,
Where deontological norm(s) DN_i are applicable,
We *can* perform action A
{Which has an effect E}
Because action A is permitted under some norm DN_x

As illustrated in Box **??**, this scheme formalises how an agent may justify an action on the grounds of permissibility rather than obligation or prohibition. Due to the nature of permission in deontological reasoning, typically understood as a derived status rather than a foundational one ($P(A) \equiv \neg O(\neg A)$), this type of argument does not introduce a distinct normative requirement, but rather identifies a boundary condition or exception to existing norms. As such, permissions may indicate the limits of a norm's reach or introduce morally relevant exceptions that must be addressed at a higher evaluative level. Moreover, permission-based arguments often arise in contexts where exceptions to general norms must be carved out or weighed against other moral principles. As such, they rarely stand alone but instead function within broader deliberative or justificatory structures. For these reasons, I treat the *Action Scheme from Deontology (permission)* as a meta-level scheme. The full treatment of this scheme and its associated critical questions is therefore deferred to Section **??**.

5 Action Scheme Appealing to Rights

Rights-based Ethics is often treated as a fourth category of normative theory [**?**], though it may also be seen as a subset of Deontology (see also Section **??**). Rights-based Ethics emerged from the three main normative theories and it focuses on establishing universally accepted standards (e.g., the right not to be enslaved). A core tenet of rights-based ethics is the correlativity of rights T and duties DN. Rights are understood as claims individuals hold against others to be treated in certain ways. To say "X has a right" implies that someone else has a duty toward X. I present the *Action Scheme Appealing to Rights* separately in order to emphasise its distinct character, but it can also be collapsed under the *Action Scheme from Deontology*. For this reason, I do not present its development in detail again, since its structure closely parallels that of the deontological action scheme.

The *Pro-choice vs Pro-life* debate is primarily framed around arguments that appeal to rights, focusing on the promotion of right $T =$ *the right to bodily autonomy* vs the violation of right $T' =$ *the right to life*. The first argument that appears in the *Kialo* debate is: "Pregnant people should have the right to choose abortion". I introduce, thus, the *Action Scheme Appealing to Rights* in Box **??**.

Box 1 Action Scheme from Rights-based Ethics

In the current circumstances R
The established right T is at stake
We should perform action A
Which has effect E
Because action A promotes or protects T

A representation of the scheme's premises in *ASPIC*$^+$ is:
f_1 : *We are in circumstances R.*
f_2 : *In R, the established right T is at stake.*
f_3 : *Action A protects/promotes right T.*
f_4 : *In R, action A has effect E.*
A formal example argument following this scheme then would be the following.

R : *Pregnant individual does not wish to carry to term*
T : *Right to bodily autonomy*
A : *Legalise abortion*
E : *The individual is not forced to remain pregnant*

A_t: "People have a right to bodily autonomy. We should legalise abortion to protect this right."

The defeasible rule of this scheme can be represented follows.
d_t (R, T, A, E): $Circumstances(R)$, $Applicable\,(R, T)$, $Aligned\,(T, A)$, $Effect(R, A, E) \Rightarrow ShouldPerform(R, A)$
So, argument A_t becomes:

$$A_t : [f_1, f_2, f_3, f_4] \Rightarrow_{d_t} ShouldPerform(R, A)$$

5.1 Critical Questions

The critical questions for this scheme are adapted from *Action Scheme from Deontology (obligation)* and are presented below.

1. Circumstances & Feasibility
    - Are circumstances R true?
    - Are circumstances R possible?
    - Are circumstances R as stated complete?
    - Is action A possible in circumstances R?
2. Relevance & Link
    - Is right T at stake in the current circumstances R?
    - Does action A protect/promote T (unambiguously) in R?
    - Does performing action A in R have a side effect E' (dictates A') that violates T?

- Are there alternative actions A' that fulfil the requirements under T more effectively or appropriately, in R?

3. Right T
   - Is right T a universally established right in R?
   - In circumstances R, is the right T subject to any recognised limitations?
4. Conflicts with T'
   - In circumstances R, does performing action A, to protect/promote T, violate some other right T'?
   - In circumstances R, does performing action A entail some other action A' that violates some T'?
   - Does performing action A also promote some other right T'?
   - Does performing action A preclude some other action A' that would promote some other right T'?
5. Other Ethical Frameworks
   - Is performing action A in conflict with moral virtue MV?
   - Does protecting T lead on balance to positive consequences?
   - Are the effects E of action A in circumstances R on balance positive consequences?
   - Does action A violate some deontological norm DN?

6 Action Scheme from Consequentialism

In contrast to deontological theories, consequentialist theories hold that the moral rightness of an action is solely based on its consequences. This reflects the basic principle that what is best or right is whatever leads to the best outcome in the future. For example, if more benefit than harm is brought about by an action, that action is good. As explained in Section **??**, there exist many subtypes of Consequentialism. Utilitarianism focuses on general well-being and Classic utilitarianism reduces all morally relevant factors to consequences, but there are many more variations of it, which are differentiated based on the measure used for well-being. For instance, *hedonistic utilitarianism* uses pleasure as the measure of well-being, *preference list utilitarianism* focuses on maximising the preferences of individuals, and *objective list utilitarianism* focuses on a set of objective goods that need to be maximised.

For the development of this scheme, Consequentialism is used as a general framework (i.e., an overarching scheme) on which specifications can be generated. I begin its development, again, by presenting some example arguments from *Kialo*. In the abortion debate, multiple arguments bypass the principles involved and focus on the consequences of an abortion ban.

> B: "When abortion is banned, many women who do not want a child seek out illegal abortions"
>
> B_a: "When abortion is banned, in practice this means that only women who can afford it can access safe abortion, since it is more expensive when illegal"

To encompass such arguments, I develop a scheme that captures reasoning grounded in Consequentialism. For this, I adapt the *argument from action* into *Action Scheme from Consequentialism* as follows. The scheme is presented in Box **??**.

- Variables

  I explicitly incorporate variables into the scheme's CQs, allowing for richer, context-sensitive CQs that are easier to formalise logically.

- Outcome focus

  The justification for acting hinges on the predicted *consequences* C_i of the proposed action A.

- Forward-looking motivation

  Action A is recommended because it is expected to move the world from the current circumstances R to new circumstances S that are explicitly better. This re-introduces the "new-circumstances" component of the original *argument from action*, but now the improvement of S is the decisive motive.

- Value metric left open

  The judgment that the consequences are *good* is judged by a metric μ that is kept generic: different ways of defining or aggregating "good" (e.g., net utility, preference satisfaction, welfare points) yield different species of Consequentialism. This placeholder invites a rich set of critical questions about:

  - how "good" is operationalised,
  - whose interests are counted, and
  - whether the comparison between R and S is well supported.

The original *argument from action* scheme already shares a broadly consequentialist structure, as it involves assessing the outcomes of actions and makes references to new states of affairs. However, the motive for the action is rooted in a goal that promotes a value and the outcomes are evaluated based on them, which is not explicitly comparative. *Action Scheme from Consequentialism* follows a similar pattern but evaluates the consequences of actions and new states of affairs according to how they score under a formalised evaluative metric μ (comparing the current circumstances with the ones that the action will bring). In Box **??**, I present a *generalised* consequentialist scheme, where different definitions of good (i.e., different choices of μ) generate different forms of Consequentialism (and, hence, different sub-schemes). For example, μ might represent utility, preference satisfaction, wellbeing, fairness, or other criteria, depending on the specific consequentialist sub-theory.

Box 1 Action Scheme from Consequentialism

In the current circumstances R
We should perform action A[a]
Which has consequences C_i
Leading to new circumstances S
Because the (good) evaluative metric μ yields $\mu(S) > \mu(R)$

[a]The scheme can also be applied in the negative direction: we should not perform action A if $\mu(S) < \mu(R)$

Below I present the scheme's premises and defeasible rule in *ASPIC*$^+$.
f_1 : *We are in circumstances* R.
f_2 : *Action* A *has consequences* C.

f_3 : *Performing action A, with consequences C, will lead to new circumstances S.*
f_4 : *There exists metric μ for evaluating the goodness of circumstances.*
f_5 : *Circumstances S are better than R according to metric μ: $\mu(R) = x$, $\mu(S) = y$ and $y > x$.*

d_c (R, A, C, S, μ): $Circumstances(R)$, $HasConsequences(A, C)$, $LeadsTo(A, C, S)$, $Metric(\mu)$, $IsBetter(S, R, \mu) \Rightarrow ShouldPerform(R, A)$

For completeness, I also include the complementary rule for when the outcome is not better:

d_{c-} (R, A, C, S, μ): $Circumstances(R)$, $HasConsequences(A, C)$, $LeadsTo(A, C, S)$, $Metric(\mu)$, $IsBetter(R, S, \mu) \Rightarrow ShouldNotPerform(R, A)$

So, a consequentialist argument A_c becomes:

$$A_c : [f_1, f_2, f_3, f_4, f_5] \Rightarrow_{d_c} ShouldPerform(R, A)$$

Below I present a formal example argument that instantiates the *Action Scheme from Consequentialism*, by adopting the classical utilitarian metric of *well-being* [**?**].

$R =$ *Economic Deprivation,* $\mu(R) = x$
$A =$ *Ban abortion*
$C =$ *Disempowering women's control of their reproductive capacities*
$S =$ *More economic deprivation,* $\mu(S) = y$
$\mu =$ *Wellbeing (by which $y < x$)*

In natural language:

> A_c : "In circumstances of economic deprivation, when abortion is banned, it has the effect of disempowering women's control of their reproductive capacities leading to further economic deprivation, which is a worse circumstance under the metric of well-being. So, we should not ban abortion."

6.1 Critical Questions

I retain many of the original critical questions from the *argument from action* (Box **??**) scheme but adapt them for consequentialist reasoning. In particular, questions that originally inquired about the promoted goal and value are reformulated to refer to comparisons under μ. These CQs include:
CQ1: Are the believed circumstances true?
CQ2: Assuming the circumstances, does the action have the stated consequences?
CQ5: Are there alternative ways of realising the same consequences?
CQ8: Does doing the action have a side effect which demotes the value (/ decreases μ)?
CQ11: Does doing the action preclude some other action which would promote some other value (/ would result in R' being better than S according to μ)?
CQ12: Are the circumstances as described possible?
CQ13: Is the action possible?
CQ14: Are the consequences as described possible?

I now group these critical questions similarly to the previous schemes. I also adopt their enhanced formulations introduced earlier. For example, CQ13 *Is the action possible?* becomes: *In circumstances R, is performing action A possible?*.

Group 1: Circumstances and Feasibility In the first group, I again include CQs concerning the truthfulness, completeness, and feasibility of circumstances R, along with the proposed action within the given circumstances, as described in the previous sections (corresponding to CQ1, 12, 13, from the original *argument from action* scheme). Since the consequentialist scheme includes new circumstances S as well, the plausibility inquiry now applies to both the initial circumstances R, but also to these resulting circumstances S. Therefore, I add to the first group the following question.

Are the circumstances S as described possible?

For example, in argument A_c, a counter-argument could contend that it is impossible for the economic situation to deteriorate further, perhaps using an *argument from expert opinion* that the economy has already reached its lowest point.

Also, I add CQ14 *Are the consequences as described possible?* and augment it as follows.

Are the consequences C as described possible?

Group 2: Link between A, C, S Next, I introduce CQs targetting the link between action A and its stated consequences C. I include the original CQ2 *Assuming the circumstances, does the action have the stated consequences?* as

In circumstances R, does action A have consequences C?

That reflects a common challenge to arguments that rely on consequence, prompting the interlocutor for evidence supporting that consequences C result from action A in R. For instance, let us formulate the aforementioned argument B from *Kialo* as an *Action Scheme from Consequentialism* B : "We should not ban abortions because, as a consequence, women with unwanted pregnancies will seek illegal alternatives, thereby leading to increased risks to women's life". A counter-argument could claim that there are measures in place that will prevent illegal abortions, challenging that in the current circumstances the action will lead to the stated consequences.

I introduce two additional CQs, derived from the Walton's *argument from consequences* (see Section **??**) that further inquire the link between the action and the stated consequences.

What evidence supports the claim that action A has consequences C and is it sufficient to support the strength of the claim adequately?

and

How strong is the likelihood that consequences C will occur in circumstances R?

These probe the empirical basis for assuming that in the stated circumstances R, A will have consequences C, effectively challenging premise f_2. In the context of the abortion debate, consider the argument in favour of legalising abortion on the grounds that it enhances the health of pregnant individuals. The former CQ necessitates a consideration of the evidence, such as research studies or historical examples, which substantiate that changes in abortion laws lead to tangible improvements in the health outcomes of pregnant people. A counter-argument to B, following the aforementioned CQ, could then respond: "According to study X, in countries where abortion is banned, the mortality rates of pregnant women do not increase".

Once evidence has been provided, the latter CQ additionally challenges whether the likelihood of the established causal link is substantial. This mirrors a common problem in consequentialism: epistemic uncertainty [**?**]. Predicting long-run, system-wide totals is notoriously error-prone and small modelling errors

can flip the ranking of states (e.g., climate-economic models vary wildly under different damage functions and discount rates). This is further addressed in Group 3 of critical questions.

Lastly, I include a CQ that inquires whether consequences C actually lead to the claimed circumstances S, pointing to undermining arguments that challenge premise f_3.

How likely is it that the stated consequences C of action A lead to circumstances S?

This question examines the causal link between the identified consequences and the resulting state of affairs, checking for logical fallacies or unjustified inferential leaps in the argument. For instance, let us consider an argument that advocates for action $A = TaxIncrease$ because it will have consequences $C =$ *generating more public revenue*, leading to circumstances $S =$ *universal free public services*. A counter-argument following this CQ could point out that, while increased taxation may indeed generate additional public revenue, the leap to assuming that this will result in universal free public services overlooks numerous intermediate steps and potential obstacles and the link is unlikely.

Group 3: Consequences C and Metric μ Once the link between action, its consequences, and the new state of affairs is established, further inquiry is needed to assess the desirability of the new, 'good' circumstances S that consequences C lead to. As previously explained, this is one of the main gaps in Walton's *argument from consequences*. There, the premise accepts the consequence's value (as good or bad) by default, and the critical questions never challenge it. However, in real debates, participants often dispute not only whether a consequence will occur, but also whether it is genuinely desirable (i.e., good).

The term 'good' requires specific qualification; it cannot be applied universally without clarifying for whom or according to what standard an outcome is good (discussion about various sub-types of Consequentialism in Section **??** illustrates how differing notions of good can affect evaluation of actions). This is a nuance that consequentialist arguments are often criticised for overlooking, assuming a universal standard of 'total good' without such distinctions [**?**]. To illustrate this, consider a general, health-related example: while improving health might seem self-evidently positive, a thorough argument would still need to specify the evaluative criteria (e.g., whether health is good because it increases life expectancy, reduces suffering, or enhances autonomy). Similarly, for argument B_t: "We should raise taxes to enhance public services", in a tax policy debate; after demonstrating that raising taxes indeed results in public service improvements, the debate moves to why this outcome is considered positive. The discussion might establish and scrutinise the social criteria that deem public service expansion beneficial. For instance, does it contribute to overall well-being?

Formalising such challenges requires addressing metric μ. As mentioned before, different types of Consequentialism adopt different metrics, defining goodness in distinct ways. I introduce a general CQ that asks whether the metric appropriately captures all the relevant considerations.

Are $\mu(R)$ and $\mu(S)$ calculated accurately?

For example, Classical Utilitarianism uses a total-sum metric of utility, aiming to "maximise total well-being for everyone (impartially)" [**?**]. Disagreements may, then, arise over the accuracy of numerical estimates (e.g., utilities, probabilities, weights) and whether well-being can be meaningfully measured across individuals at all. Arguments that are based on the premise that $\mu(R)= x$ and $\mu(S)= y$ can be countered with arguments claiming that $\mu(R)= x'$, and $x' > y$, or $\mu(S)= y'$, and $x > y'$, etc. Such challenges point to undermining attacks towards premise f_5. For example, a counter-argument against A_c could claim: "You are only counting women's economic well-being, you should also take the well-being of the fetus into account. When you factor in the positive utility of the new life, the new circumstances are not worse than the current ones". It becomes apparent that even this (seemingly intuitive) metric requires clarification and invites various challenges. Who counts as "everyone"? Should we include potential people (future generations, unborn individuals)? If yes,

should they be counted equally with the current people? In the aforementioned counter-argument, this seems pivotal. Of course, even if we do accept that $\mu(S)$ is not y as we initially claimed (i.e., in the aforementioned example, that the welfare of the future child should also be calculated), we still need to conclude whether this changes which state is the better one.

Let us take the following formal representation of this metric μ for the new circumstances.

$$\mu_(S) \;=\; \sum_{i=1}^{N_S} u_i(S),$$

$$\begin{aligned} S &= \text{the state of affairs being evaluated,} \\ N_S &= \text{the number of sentient beings that exist in } S, \\ u_i(S) &= \text{the net balance of well-being experienced by individual } i \text{ in } S. \end{aligned}$$

Using this CQ, three key components of the metric can be challenged (1) who/what gets included in the sum (N_S); (2) how much utility each entity has $u_i(S)$; (3) the aggregation itself, i.e. how are these utilities summarised. In the aforementioned example, N_s is disputed, debating whether the fetus should be included in it. Further challenges might include what utility to attribute to it, since calculating individual utility can be challenging in practice, as it requires measuring subjective experiences. Lastly, one might say that probabilist weights should be included when summing the utilities, potentially arguing that a future person should be attributed a lower weight. As each variant of consequentialism adopts a different μ metric-specific questions going into the details of the calculations should be handled within the respective sub-scheme of Consequentialism. For this reason, I do not attempt an exhaustive list of such questions. Instead, I highlight that disputes over μ are a critical part of real-world deliberation and offer an overview of such critical questions.

I create two general CQs that challenge the universality of the metric and suggesting that different standards might lead to contrasting evaluations of the same outcomes. For example, an alternative μ' might prioritise average well-being instead of total. Arguments regarding such comparisons between metrics belong at the meta-level (see Section **??**). Here, I introduce first the following question to address the acceptability of the adopted metric.

Is μ a widely accepted metric?

This CQ challenges premise f_4 by questioning the adoption of the metric *per se*. For instance, one might contest the use of a utilitarian metric altogether. A prominent example is Parfit's *repugnant conclusion*, arguing that total-sum utilitarianism can imply that adding vast numbers of lives barely worth living is better than a smaller, flourishing population [**?**]. An example argument responding to the aforementioned CQ is "Your consequentialist argument relies on a total utilitarian metric that is fundamentally flawed and should be not be trusted as it leads to counterintuitive conclusions that violate our basic moral intuitions (e.g., a world with millions of people with lives barely worth living would be preferred over a world of thousands where people flourish)". Successful challenges of this type might shift the debate from consequence evaluation to framework selection, moving the argument to the meta-level where different moral theories themselves become the subject of dispute.

Building on this, I additionally introduce a second CQ that directly points to the comparison of two evaluative metrics.

Is there an alternative metric μ' under which circumstances S are not assessed as better than R ($\mu'(R) \not< \mu'(S)$)?

For example, consider an immigration policy debate where circumstances R involve a country with 10 million citizens enjoying an average well-being of 8 utils per person, yielding a total well-being of 80 million utils.[12] Action A to open borders would have the consequence of 5 million immigrants coming into the country. This would result in new circumstances S with 15 million people, but due to resource strain and increased competition, the average well-being drops to 6 utils per person. Under total utilitarianism, $\mu(R) = 80$ million utils and $\mu(S) = 90$ million utils, thus action A should be performed since $\mu(R) < \mu(S)$. However, under the alternative metric μ' based on average utilitarianism, $\mu(R) = 8$ utils pp and $\mu(S) = 6$ utils pp, thus action A should not be performed since $\mu'(S) < \mu'(R)$. Again, the resolution of which metric is more appropriate would take place on the meta level.

After ensuring the right metric has been adopted, one can consider whether all relevant consequences have been included and appropriately articulated (and, hence, calculated). I introduce the following CQ to address this.

Are there additional consequences C' of action A that have not yet been considered?

This question challenges the completeness of the consequence set by suggesting that the original argument may have overlooked significant effects of the proposed action. For instance, in the abortion debate, an argument advocating $A =$ *widespread access to abortion* might focus primarily on consequences C_i, including health and autonomy benefits, resulting in S that represent a safer society. However, a counter-argument employing this CQ could point to additional consequence C' such as potential demographic changes (decreased birth rates), which might subsequently result in S' that include labor market shortages and long-term economic challenges. Such challenges highlight the complexity of consequentialist reasoning, where the identification and articulation of all relevant consequences becomes crucial for accurate evaluation. The inclusion of previously unconsidered consequences C' may significantly alter the calculation of $\mu(S)$.

This example touches upon another significant factor in the consequentialist calculation that was hinted upon previously: the temporal aspect in the prediction of the new circumstances. I introduce a special case of the CQ above to reflect the significant criticism of Consequentialism regarding the distinction between expected consequences and long-term consequences ($C' = C_{LT}$).

Have potential long-term consequences C_{LT} been considered?

Consider a debate about AI-assisted medical diagnosis where an argument advocates for $A =$ "We should allow AI systems to override physician recommendations when the AI's diagnostic confidence exceeds 95%" based on the immediate consequences that this will save lives by catching cases where human doctors make errors, leading to circumstances S representing improved diagnostic accuracy and better patient outcomes. However, a counter-argument responds: "What about the long-term consequences? Following such a principle could lead to doctors becoming overly dependent on AI systems, losing their diagnostic expertise over time", leading to circumsntaces S' where patients lose trust in human medical judgment and complex ethical decisions about patient care are increasingly delegated to algorithms.

If the aforementioned CQ reveals C', additional challenges arise. First, about the way these consequences influence our initial calculations.

How do these consequences C' influence the valuations of metric μ for S?

[12] For the purposes of this example, I assume utils to be a commonly accepted unit of measurement of well-being, between the two frameworks.

I do not elaborate extensively on the possible sub-questions, as this again gets into the specifics of the metric. However, it is important to emphasise that when new elements come into play, a point of discussion should be about the manner they fit in. Particularly when including long-term consequences, temporal weighting becomes crucial: does a hypothetical person who will live in a thousand years carry the same weight as a person alive today? How far into the future should S extend?

The previous example points to an important consideration that emerges from the rule consequentialist challenge, which addresses the universalisation of actions and their precedent-setting effects. This approach evaluates not merely the direct consequences of a specific action, but examines what would happen if everyone followed the rule that would justify that action.[13] This distinction proves particularly relevant in addressing one of the most prominent objections to Act Consequentialism: the transplant case. Critics argue that if saving five lives justifies sacrificing one, then it would be morally permissible for a physician to kill a healthy patient to harvest their organs for five dying patients [**?**]. The rule consequentialist response maintains that while the immediate consequences might appear beneficial (five lives saved versus one lost), the universal application of such a principle would be catastrophic, because if healthcare systems operated under the rule, that would undermine trust in medical institutions, leading to widespread avoidance of healthcare and ultimately far more deaths [**?**].

In an ethical discourse, precedent effects might be brought up as additional consequences C'. Consider a court case where a prosecutor argues for A : "Convict the defendant for hate speech posted on social media", based on the (direct) consequences C that include protecting vulnerable communities and deterring similar harmful behavior, resulting in circumstances S representing a safer society for minorities. However, a defense attorney might counter with a rule consequentialist precedent argument: "While the immediate consequences of conviction might seem beneficial, consider the precedent this ruling would establish. If we allow convictions for this type of online speech, what happens when this legal principle is universally applied? The long-term consequences (C_{LT}) include a systematic erosion of free speech protections. Future prosecutors will cite this case to justify increasingly broad interpretations of what constitutes incitement".

These cases of long-term consequences and precedents create additional complexity for the valuations of metric μ, extending the temporal horizon significantly. This naturally leads to questions about how far into the future our consequentialist calculations should extend. The temporal factor that I mentioned in the previous group ("Is the future child part of the valuation of the new circumstances?") becomes relevant once again. The decision of how far in the future the new circumstances S should extend brings us to longterminism [**?**], which argues that what most matters morally is ensuring that the long-term future goes well, given the vast number of potential future lives and their cumulative well-being. The temporal weighting then becomes crucial: does a hypothetical person that will live in a thousand years have the same weight as an alive person now? What are the probabilities that the Earth will remain habitable until then? How shall this be represented in new circumstances S? This raises classic dilemmas about discount rates and relates to the cluelessness problem [**?**], where the long-term consequences of our actions are so uncertain that confident evaluation becomes difficult.

Finally, the last CQ of this group is an augmentation of CQ8 *Does doing the action have some unstated consequence C' which decreases μ' valuation of S?*. It asks whether, once the new consequences are incorporated and the valuation of circumstances is updated, the resulting circumstances S remain better than the current circumstances R. This points to undermining attacks that challenge premise f_5.

Is it the case that $\mu(S) > \mu(R)$ still holds if S includes C'?

[13] The rule consequentialist approach connects to Kantian universalisation principles while maintaining consequentialist evaluation criteria.

Group 4: Alternative Actions A' Lastly, after considering conflicts arising from competing consequences, I turn to conflicts stemming from alternative actions. In this case, the issue is not additional side effects, but whether there exists a different action that could achieve the same desirable outcomes, potentially more effectively or with fewer costs. I first adapt CQ5 ("Are there alternative ways of realising the same consequences?") as follows. Responding to both these critical questions leads to rebutting arguments.

Is there an alternative action A' that realises consequences C?

and

Is there an alternative action A' that can bring about S?

Lastly, I refine the discussion further. The debate often pivots around competing actions. Consider a healthcare resource allocation scenario. One possible action A is to approve an expensive cancer treatment that extends the lives of terminal patients by an average of three months. Alternatively, action A' would allocate the same budget to preventative care measures (e.g., vaccinations) that could save a greater total number of life-years across the population. Both A and A' aim for consequences $C =$ "reduce mortality", but they achieve this through different pathways with distinct trade-offs. Moreover, performing both actions is impossible (thus, they are mutually exclusive).[14] I introduce the following CQ, adaptation from CQ11 ("Does doing the action preclude some other action which would result in R' better than S by μ?"), which points to rebutting arguments.

Does doing action A preclude some other action A' that might lead to better circumstances S', according to μ ($\mu(S) < \mu(S')$)?

Group 5: Other Ethical Frameworks Lastly, analogously to the previous schemes, I incorporate CQs that examine compatibility with other ethical frameworks. As before, I note that clashes between moral frameworks often shift the discourse to the meta level, where the appropriateness of different frameworks for the given circumstances becomes the subject of debate (see Section **??**).

First, I introduce a CQ that examines incompatibility with deontological norms. As discussed in the previous section, debates about abortion often culminate in a clash between consequentialist and deontological reasoning. A consequentialist argument might state "We should legalise abortion because it leads to better health outcomes and reduced maternal mortality, resulting in overall greater societal well-being". This can be challenged by a deontological argument asserting that action A (legalising abortion) violates the deontological prohibitory norm $DN_p =$ "One should not kill an innocent life", asserting that regardless of the positive outcomes, the action itself is inherently wrong and should not be performed. The corresponding CQ is formulated as follows.

In circumstances R, does action A violate a deontological norm DN?

Second, I consider challenges from Virtue Ethics. Let us examine the trolley problem, particularly the "fat man" variant [**?**], where a person faces the dilemma of either allowing a trolley to kill five people or pushing a large person off a bridge to stop the trolley. In these circumstances R, a consequentialist argument might

[14] This case could formally be modelled as an additional consequence C', where $C' =$ " not performing action A'". However, I treat it explicitly as an alternative action A' for clarity and practical reasons. First, identifying A' involves a distinct cognitive step from listing side effects of A (it requires recognising different branches of the decision tree, not just consequences of a single path), this is helpful when making comparisons on the meta-level. Second, opportunity costs are a common source of reasoning failure in real-world debates, and highlighting them with a dedicated CQ ensures they are not overlooked. Finally, in decision-support and argument-mapping systems, alternative actions are typically represented as separate nodes from consequences.

advocate for $A =$ *push the large person off the bridge to stop the trolley* because this saves five lives at the cost of one (C), maximising overall well-being. However, a challenge from Virtue Ethics would contend that in the described circumstances R, a virtuous person would not perform action A (pushing someone to their death) as this would require viewing the big person as a heavy object, merely as an instrument rather than as a fellow human being deserving of moral respect. The corresponding CQ is formulated as follows.

In circumstances R, is performing action A in conflict with moral virtue MV?

Lastly, I address challenges based on rights violations. Consider again the scenario of the captured terrorist. A consequentialist argument could advocate for $A =$ *torture this terrorist* to achieve $C =$ *extract information about an imminent attack that will kill thousands of innocent people*, thereby saving thousands of lives, maximising overall wellbeing. An *Action Scheme Appealing to Rights* could rebut this argument, arguing not to perform the proposed action A as A violates the (universally) established human right $T =$*freedom from torture*. The corresponding CQ is formulated as follows.

In circumstances R, does action A violate some universally established right T?

Summary Below, I summarise all the CQs for *Action Scheme from Consequentialism*.

1. Circumstances & Feasibility
   - Are circumstances R true?
   - Are circumstances R possible?
   - Are circumstances R as stated complete?
   - Are circumstances S as stated possible?
   - Is action A possible?
   - Are the consequences C as described possible?
2. Link between A, C, S
   - In circumstances R, does action A have consequences C?
   - How strong is the likelihood that consequences C will occur in circumstances R?
   - What evidence supports the claim that action A has consequences C and is it sufficient to support the strength of the claim adequately?
   - How likely is that the stated consequences C of action A lead to the new circumstances S?
3. Consequences C & Metric μ
   - Are $\mu(R)$ and $\mu(S)$ calculated correctly?
     (a) Who/what should be included in the calculation?
     (b) How should utilities be assigned?
     (c) What weights are being distributed?
   - Is μ a widely accepted metric?
   - Is there an alternative metric μ' under which circumstances S are not assessed as better than R ($\mu'(R) \not< \mu'(S)$)

- Are there additional consequences C' of action A that have not yet been considered?
  - Have potential long-term consequences C_{LT} been considered?
- How do these consequences C' influence the valuations of metric μ for S?
  - Is it the case that $\mu(S) > \mu(R)$ still holds if S includes C'?

4. Alternative Actions A'

   - Is there an alternative action A' that realises consequences C?
   - Is there an alternative action A' that can bring about S?
   - Does doing action A preclude some other action A' that might lead to better circumstances S', according to μ ($\mu(S) < \mu(S')$)?

5. Other Ethical Frameworks

   - In circumstances R, does action A violate a deontological norm DN?
   - In circumstances R, is performing action A in conflict with a moral virtue MV?
   - In circumstances R, does action A violate some universally established right T?

7 Meta Level: Intra-Theoretical Comparisons

Conflicting arguments are a common phenomenon in ethical dilemmas, arising both within a single theory and across different theories, as competing recommendations often privilege one 'horn' of the dilemma over the other. Adopting a specific ordering or preference effectively prioritises one side, and the reasoning that governs how and when such preferences are asserted operates at a meta-level. To capture this, I introduce a new category of argument schemes that is not currently covered by existing scheme taxonomies: meta-level schemes. During the development of the new argument schemes for ethical reasoning in the previous section, we repeatedly encountered situations where opposing arguments arose from conflicts; conflicts between duties, rights, consequences, etc. Such clashes are pervasive in ethical dilemmas (hence their very nature as dilemmas) and are central to ethical debates. For example, the *Pro-choice vs Pro-life* debate highlights conflicts such as the tension between the right to bodily autonomy and the right to life. Conflicts of this kind can occur both within arguments that draw on the same ethical theory (as in the example above, where both sides appeal to rights-based reasoning) and across different ethical theories.

As discussed in Section **??**, arguments may also express preferences over other arguments, operating on a meta level. Related work has demonstrated the effectiveness of meta-level frameworks—for instance, Kokciyan et al. [**?**] apply such frameworks to support medical decision-making. Building on this, I introduce *meta-schemes*, designed specifically to express preferences and enable commentary on the reasoning of the object-level reasoning process itself. Meta-schemes support deliberation about which arguments should take precedence in a given context by enabling reasoning about preferences across arguments, ethical frameworks, and evaluative criteria. They help determine which action should be preferred, especially when object-level arguments are in conflict. They support intra-theoretical comparisons, providing a systematic way to deliberate over which ethical considerations should take precedence when clashes arise. This resolves symmetric attacks by converting bidirectional conflicts into unidirectional ones, in much the same way that preferences are incorporated into structured argumentation frameworks (see Section **??**).

It is important to note that the meta-schemes presented in this section do not constitute an exhaustive taxonomy. Rather, they represent types of meta reasoning identified through the descriptive study detailed in

Section **??** and from the systematic analysis of conflicts between our newly crafted object-level arguments the previous sections.

7.1 Virtue Ethics Orderings

As demonstrated in Section **??**, mutually attacking arguments can emerge that both appeal to Virtue Ethics, either stemming from appeal to different virtues or arguing about which action exemplifies a certain virtue better. For instance, within the abortion-related arguments, we previously encountered the following symmetrically attacking arguments.

> Arg_1: "Choosing to raise a child when one is not ready is an irresponsible decision"
> Arg_2: "A person is actually responsible when they recognise they cannot raise a child and choose not to"

The meta-scheme addresses such conflicts by articulating a context-dependent argument that expresses prioritisation. Specifically, I develop two schemes: one to model conflicts arising from appeals to the same virtue, and another to capture conflicts between appeals to different virtues. The construction of these schemes follows the steps outlined below.

- I create a variable Arg to represent the object-level schemes from Virtue Ethics. This entails that their format is embedded in the meta-schemes without explicitly repeating the schemes (so, the variables of the object level scheme are included in the meta-scheme).
- I introduce a preference ordering, which in natural language incorporates the justification for the preference. The meta-scheme does not explicitly spell out the reason for the preference; rather, the justification is embedded within the ordering itself.

In Box **??**, I introduce the meta-scheme for deliberation over competing actions that both exemplify the same virtue, as the ones in the example above. The two symmetrically attacking arguments instantiating the *Action Scheme from Virtue Ethics* are :
Arg_1: $Circumstances(R)$, $ExemplifiesVirtue(R, A, VA, MV)$, $Effect(R, A, E)$, $Goal(R, A, G) \Rightarrow ShouldPerform(R, A)$ and
Arg_2: $Circumstances(R)$, $ExemplifiesVirtue(R, A', VA, MV)$, $Effect(R, A', E')$, $Goal(R, A', G') \Rightarrow ShouldPerform(R, A')$[15]

In this case, the preference relation of the meta-scheme $>_{MV}$ expresses that an argument exemplifies the virtue MV better. In natural language, this can be articulated as an argument establishing the prevalence of Arg_2 over Arg_1; for example, by claiming that Arg_2 better exemplifies MV because it demonstrates genuine self-knowledge and an honest assessment of one's capabilities. Thus, the preference relation ($Arg_1 >_{MV} Arg_2$) can incorporate the rationale for why one action is judged superior in terms of virtue MV. How well this rationale is substantiated is evaluated, in turn, through the associated critical questions.

Then, in Box **??**, I introduce a meta-scheme to deliberate over mutually attacking arguments that appeal to different virtues. These cases address the conflicts that arise from Group 4 of the critical questions for the Virtue Ethics scheme (Section **??**), which encompasses situations where arguments either advocate for different, mutually exclusive actions based on competing virtues, or where arguments invoke different virtues to support opposing positions regarding the same action. Similarly to the previous meta-scheme, the preference relation $\prec$ between virtues encompasses the natural language rationale.

[15] It can also be the case that the goal is the same, G.

Box 1 Meta-Scheme for Virtue Ethics (same virtue)

In the current circumstances R
Mutually attacking arguments Arg_1 and Arg_2 advocate for actions A_1 and A_2[a]
Which both exemplify virtue MV
We should perform action A_1
Because A_1 exemplifies virtue MV better ($A_1 >_{MV} A_2$)

[a]This can also capture cases of direct opposition, where the two competing arguments advocate for A and $\neg A$ respectively, with $A_1 = A$, $A_2 = \neg A$.

Box 2 Meta-Scheme for Virtue Ethics (Generalised)

In the current circumstances R
Arg_1 advocates for action A_1 because of MV_1
Arg_2 advocates for action A_2 because of MV_2[a]
We should perform action A_1
Because MV_1 is preferred to MV_2 ($MV_2 \prec MV_1$)

[a]This can also capture cases of direct opposition, where the two competing arguments advocate for A and $\neg A$ respectively, with $A_2 = \neg A1$. In this case, 'because of MV_2' is interpreted as 'because A_1 is in conflict with MV_2'.

Critical Questions As the object level arguments are embedded in the meta-schemes, the critical questions applicable to object-level *Action Scheme from Virtue Ethics* remain relevant at the meta-level. In natural dialogue, it is common for interlocutors to shift between meta-level virtue comparisons and object-level challenges. For instance, consider the following dialogue,

- I should report my colleague's misconduct because that is what a virtuous person would do, it shows courage.
- You should talk to them privately first because that shows true courage; facing them directly.
- No, reporting it exemplifies courage more; it takes real bravery to risk workplace relationships.
- Well, reporting them will not change anything anyway.

This dialogue involves *Action Schemes from Virtue Ethics* Arg_1, Arg_2, where $R =$ misconduct, $A_1 =$ report, $A_2 =$ not report, $MV =$ courage. The dialogue can be analysed as follows.

-Arg_1 concluding $ShouldPerform(R, A_1)$ appealing to MV
-Arg_2 concluding $ShouldPerform(R, A_1)$ appealing to MV
-Meta-scheme for Virtue Ethics concluding $ShouldPerform(R, A_1)$ because of preference $A_1 >_{MV} A_2$
-Object-level argument (specifically, *argument from cause to effect*) that answers the CQ *Does action A_1 have effect E_1 in circumstances R?* (i.e., some expected effect of misconduct rectification).

Beyond these, a meta-scheme can face challenges adjacent to the meta-level comparison. In this section, I focus exclusively on challenges to the meta-argumentative reasoning and introduce the new meta-questions.

First, I question whether the proposed actions are genuinely incompatible, as some conflicts may be resolvable through sequential or joint implementation rather than prioritisation (essentially, whether this is, in fact, a bidirectional attack). For instance, in a debate about responding to a natural disaster, one might argue for immediate rescue operations (exemplifying compassion) while another advocates for systematic evacuation planning (exemplifying prudence). However, these actions may not be mutually exclusive—rescue operations could be implemented immediately while evacuation planning proceeds in parallel.

Are A_1 and A_2 strictly incompatible in R, or can they be jointly or sequentially satisfied?

Second, I introduce a CQ to assess the adequacy of the justification underpinning the ordering. For instance, in the abortion example above, one might challenge whether the claim that abortion reflects greater self-awareness constitutes a sufficient reason to conclude that abortion exemplifies responsibility to a higher degree. Similarly, in the workplace misconduct dialogue, one could claim that the link between the justification and the preference is not adequately established (e.g., why should risking workplace relationships be considered a greater display of bravery?).

How well is the ordering $MV_1 \succ MV_2$ / $A_1 >_{MV} A_2$ justified? What grounds justify the preference ordering?

Subsequently, I introduce critical questions that target specific ways in which a proposed ordering may be challenged. First, an ordering may be presented as a general preference that does not necessarily hold in particular circumstances R because of the degree to which an action exemplifies or is in conflict with a virtue. Consider a scenario where one must choose between telling a harsh truth (exemplifying honesty) or offering comfort through gentle reassurance (exemplifying compassion). One argument favouring the first action can be based on a preference like "honesty always before compassion". However, in the given circumstances, this preference may not apply: if truth-telling severely conflicts with compassion while reassurance only mildly compromises honesty, this asymmetry of conflict can outweigh a general preference ordering between virtues. This motivates the following CQ:

Is action A_1 in conflict with MV_2 to the same degree that A_2 is in conflict with MV_1?

Analogously, different virtues may be invoked to support competing actions, and the degree to which each action exemplifies its virtue can vary. For instance, the virtue $MV =$ *responsibility* might support $A_1 =$ *abolish abortion*, whereas $MV' =$ *compassion* might support $A_2 =$ *legalise abortion* as a means of alleviating suffering. In such cases, the question becomes whether one action exemplifies its virtue more fully than the other. This leads to a second CQ:

Does action A_1 exemplify MV_1 to the same degree that A_2 exemplifies MV_2?

Then, I introduce CQs that focus on the acceptability of the cited ordering, which may depend on social or contextual norms. For example, in medical contexts honesty is generally preferred over compassion when delivering difficult diagnoses, whereas in personal relationships the reverse ordering is often more acceptable. Two variants of this challenge arise: one appeals to common practice, (e.g., "In medical fields, honesty should be preferred to compassion when communicating diagnoses"), while the other appeals to virtuous agents (e.g., "A virtuous person would never prioritise compassion over honesty"). These considerations motivate two CQs:

Is the ordering $MV_1 \succ MV_2$ / $A_1 >_{MV} A_2$ aligned with common practices?

Is the ordering $MV_1 \succ MV_2$ / $A_1 >_{MV} A_2$ consistent with what VAs would do?

The last consideration addresses the inherent limitations of single theoretical frameworks and questions whether this dilemma should be resolved within Virtue Ethics. This opens the possibility of inter-theoretical resolution, modelling arguments such as: "Since responsibility and compassion conflict irreconcilably here, we should employ consequentialist reasoning to determine which action produces better outcomes". This question bridges intra-theoretical meta-reasoning with inter-theoretical comparison, suggesting that framework selection itself becomes the appropriate level of analysis when internal ordering mechanisms prove inadequate. I elaborate on inter-theoretical comparison in Section **??**.

Should this virtue conflict be resolved by appealing to an alternative moral framework rather than this prioritasation?

Below the critical questions are summarised.

1. How well is the ordering $MV_1 \succ MV_2$ / $A_1 >_{MV} A_2$ justified? What grounds justify the preference ordering?
2. Are A_1 and A_2 strictly incompatible, or can they be jointly or sequentially satisfied?
3. Is action A_1 in conflict with MV_2 to the same degree that A_2 is in conflict with MV_1?
4. Is action A_1 exemplifying MV_1 to the same degree that A_2 is exemplifying MV_2?
5. Is the ordering $MV_1 \succ MV_2$ / $A_1 >_{MV} A_2$ aligned with common practices?
    - Is the ordering $MV_1 \succ MV_2$ / $A_1 >_{MV} A_2$ consistent with what VAs would do?
6. Should this virtue conflict be resolved by appealing to an alternative moral framework rather than this prioritasation?

7.2 Deontology Orderings

Using the same approach, I develop a meta-level scheme designed to adjudicate between arguments that motivate actions through appeals to deontological norms. As demonstrated in Section **??**, it is common for mutually conflicting arguments to emerge from appealing to deontological reasoning but stem from incompatible deontological norms. For instance, in the 'murder at the door' scenario presented in Section **??**, the norms DN_p ="One should not lie" and DN_o ="One must protect the innocent" point to different, incompatible actions. Two arguments instantiating *Action Scheme from Deontology*, A_1 based on DN_p and A_2 based on DN_o. Such conflicts necessitate an ordering of the norms at play if one wants to resolve the mutual attack between the arguments. To encompass such arguments, I develop meta-schemes for Deontology with the following steps.

- Creating a variable Arg to represent the object-level schemes from Deontology.
- Introducing a preference ordering, which in natural language will be citing the justification for the preference (e.g., in the aforementioned example, one could argue that DN_o takes precedence as the potential murderer has forfeited their right to honesty since they are acting immorally within this deontological framework).

First, in Box **??**, I introduce the meta-scheme to deliberate over competing actions that both cite the same norm, as in the case of the example of the factory discussed in the second group of critical questions for *Action Scheme from Deontology (obligation)* in Section **??**. This norm can be either a prohibition or an obligation. Two competing arguments based on norm DN_o = *Minimise Carbon Footprint*, one advocating for A = *Relocate Abroad*, the other for A_2 =*Install on-site Installations*.

Box 3 Meta-Scheme for Deontology (same norm)

In the current circumstances R
Mutually attacking arguments Arg_1 and Arg_2 advocate for actions A_1 and A_2[a]
Based on norm DN
We should perform action A_1
Because A_1 fulfills the requirement of DN better ($A_1 >_{DN} A_2$)

[a]This can also capture cases of direct opposition, where the two competing arguments advocate for A and $\neg A$ respectively, with $A_1 = A$, $A_2 = \neg A$.

Then, in Box **??**, I introduce a meta-scheme to deliberate over mutually attacking arguments that cite different norms. These cases address the conflicts that arise from Group 4 of the critical questions of the deontological schemes, which encompasses situations where arguments either advocate for different, mutually exclusive actions based on different norms, or where arguments invoke different norms to support opposing positions regarding the same action. These norms can be either prohibitions or obligations. Take, for example, the central normative conflict in Sophocles' *Antigone* [**?**]. On one part, Creon forbids the burial of Polyneices based on the civic prohibition DN_p = "One ought not to bury traitors"; on the other hand, Antigone appeals to divine and familial obligation DN_o = "The dead must be given rites". In circumstances R, where Polyneices, Antigone's brother, attacked his own city and died, the two norms prescribe directly opposed courses of action for Antigone, in regards to action A =*Bury Polyneices*. In the object level, this is unresolveable, as it leads to two mutually attacking arguments Arg_p, Arg_o instantiating the scheme *Action Scheme from Deontology*, with $Concl(Arg_p)$=$\neg A$ and $Concl(Arg_o)$=A . At the meta level, an argument instantiating *Meta-scheme for Deontology* could be in favour of Antigone burying her brother based on a prioritasation of the divine law over the civic one, $DN_p \prec DN_o$.

Box 4 Meta-Scheme for Deontology (Generalised)

In the current circumstances R
Arg_1 advocates for action A_1 to adhere to DN_1
Arg_2 advocates for action A_2 to adhere to DN_2[a]
We should perform action A_1
Because there is a preference relation $DN_2 \prec DN_1$

[a]This can also capture cases of direct opposition, where the two competing arguments advocate for A and $\neg A$ respectively, with $A_2 = \neg A1$.

Lastly, in Box **??**, I formulate a third meta-scheme to capture specifically the deontological norms that introduce permissions/exceptions (see also Section **??**). For example, one can imagine a counter-argument to Creon grounded in a permission norm, such as DN_x: "In the case of close kin, one is permitted to bury even

a traitor". This norm does not impose an obligation but relaxes the civic prohibition DN_p in light of close familial ties, based on the extremity of such punishment in ancient Greece.

Box 5 Meta-Scheme for Deontology (permission)

In the current circumstances R
Arg_1 advocates not to perform action A because it it prohibited under DN_p
Arg_2 advocates we can perform action A because it is permitted under DN_x
We should perform action A
Because DN_x is applicable

Critical Questions Similarly to before, the critical questions applicable to the object-level deontological arguments remain relevant at the meta level. In this section I introduce the additional meta-CQs, which follow the same logic as the virtue-based meta-CQs, but are adapted to the structure of deontological reasoning.

Firstly, I introduce again the CQ about (in)compatibility of actions.

Are A_1 and A_2 strictly incompatible, or can they be jointly or sequentially satisfied?

The second CQ focuses on the justification of the preference ordering. In the example case of Antigone, Creon prioritises holding civic order and authority above kinship. Antigone prioritises holding divine law and familial duty above human decree. One could argue that the laws of the gods are eternal and outrank any human decree, hence Antigone should bury her brother.

How well is the ordering $DN_1 \succ DN_2$ / $A_1 >_{DN} A_2$ justified? What grounds justify this preference ordering?

The third CQ is again about the degree of violation. Let us consider a case involving traffic laws. Suppose I am driving to escape someone who intends to harm me, and I reach a train crossing as a train approaches. If I accelerate, I can cross before the train and escape, but in doing so I would break the speed limit. Here, two deontological norms conflict: DN_1, a duty to obey the law, DN_2, a duty to preserve yourself. The competing arguments from deontology are: Arg_1: "You should wait for the train to adhere to DN_1" and Arg_2: "You should cross quickly to adhere to DN_2. In the described circumstances R, crossing the tracks only minimally violates DN_1 (a slight speeding infraction), whereas waiting entirely violates DN_2 (a complete failure to preserve life). The asymmetry in the degree of violation thus motivates a preference ordering between the two norms.

Does action A_1 derive from DN_1 to the same degree that A_2 derives from DN_2?

The following CQ concerns whether the cited ordering aligns with common practice. For instance, the prohibition of burial was especially severe in Greek religion, where proper funeral rites were considered a sacred obligation. To leave a corpse unburied was to condemn the soul to wander without rest, excluded from the gods of the underworld. Thus, the not burying punished his body and threatened his eternal fate and and it was really against common practices.

Is the ordering $DN_1 \succ DN_2$ / $A_1 >_{DN} A_2$ aligned with common practices?

To make this more concrete, since deontological norms are often found in institutional or legal domains where such priorities are codified, I introduce a subsequent CQ, asking specifically if there is an established or official ordering of the relevant norms. For example, consider a conflict between a contractual non-disclosure agreement (NDA) and a subpoena in a court of law. $DN_c =$ You should not disclose information (NDA) $DN_l =$ You should obey the law

In circumstances R where you are called to testify about information covered by the NDA, two mutually attacking arguments arise:

Arg_1: You must not reveal the requested information, because this is prohibited by DN_c.

Arg_2: You must disclose the requested information, because it is obliged under the legal duty DN_l to comply with the subpoena.

Someone could argue that the information should not be disclosed, since an NDA is more case-specific and thus its prohibition stronger than a general obligation to the law. However, many legal systems stipulate an official ordering: the duty to obey the law overrides the contractual duty created by a private agreement. In this case, the preference relation $DN_c \prec DN_l$ is institutionally justified. The associated meta-CQ can thus be formulated as:

> *Is there an official or legally established ordering between DN_1 and DN_2?*

The last CQ addresses again the limitations of single theoretical frameworks and questions whether this dilemma should be resolved within a deontological framework. Returing to the example of Antigone, consider an argument claiming that Antigone should not bury her brother because obeying the law is above familial duty. This argument could be countered by appealing to Virtue Ethics: rather than adjudicating between norms, the right course of action is the one a virtuous person would perform. Here, the relevant virtues MV in favor of Antigone's position include justice (giving the dead their due), piety (honouring the gods), and familial loyalty. Despite Creon also appealing to virtues such as order, the narrative of the tragedy paints him in a color that his insistence on order reflects an excessive rigidity that undermines practical wisdom. Thus, on a virtue ethical account (in combination with accrual), A would be affirmed as the preferred action.

> *Should this norm conflict be resolved by appealing to an alternative moral framework rather than this prioritisation?*

The critical questions are summarised as follows.

1. Are A_1 and A_2 strictly incompatible, or can they be jointly or sequentially satisfied?
2. How well is the ordering $DN_1 \succ DN_2$ / $A_1 >_{DN} A_2$ justified? What grounds justify this preference ordering?
3. Does action A_1 derive from DN_1 to the same degree that A_2 derives from DN_2?
4. Is the ordering $DN_1 \succ DN_2$ / $A_1 >_{DN} A_2$ aligned with common practices?
    - Is there an official or legally established ordering between DN_1 and DN_2?
5. Should this norm conflict be resolved by appealing to an alternative moral framework rather than this prioritisation?

Box 6 Meta-Scheme for Rights-based Approach (same right)

In the current circumstances R
Mutually attacking arguments Arg_1 and Arg_2 advocate for actions A_1 and A_2[a]
Based on right T
We should perform action A_1
Because A_1 protects/promotes T better ($A_1 >_T A_2$)

[a]This can also capture cases of direct opposition, where the two competing arguments advocate for A and $\neg A$ respectively, with $A_1 = A$, $A_2 = \neg A$.

Box 7 Meta-Scheme for Rights-based Approach (Generalised)

In the current circumstances R
Arg_1 advocates for action A_1 to protect to T_1
Arg_2 advocates for action A_2 to protect to T_2[a]
We should perform action A_1
Because there is a preference relation $T_2 \prec T_1$

[a]This can also capture cases of direct opposition, where the two competing arguments advocate for A and $\neg A$ respectively, with $A_2 = \neg A1$.

7.3 Rights Orderings

As previously explained, a rights-based ethical approach can be collapsed under Deontology. This section presents the meta-scheme for Rights-based Ethics, along with its critical questions, directly adapted from the meta-schemes for Deontology.

Critical Questions

1. Are A_1 and A_2 strictly incompatible, or can they be jointly or sequentially satisfied?
2. How well is the ordering $T_1 \succ T_2$ / $A_1 >_T A_2$ justified? What grounds justify the preference ordering?
3. Does action A_1 promote T_1 to the same degree that A_2 promotes T_2?
4. Does action A_1 violate T_2 to the same degree that A_2 violates T_1?
5. Is the ordering $T_1 \succ T_2$ / $A_1 >_T A_2$ aligned with common practices?
    - Are there any legally established orderings (e.g., court rulings) between T_1 and T_2?
6. Should this rights' conflict be resolved by appealing to an alternative moral framework rather than this prioritisation?

7.4 Consequentialism Orderings

In the same line of thought, the last meta-scheme for intra-theoretical comparison is created. This meta-scheme models arguments that compare arguments instantiating consequentialist schemes. Because the *Action Scheme from Consequentialism* evaluates outcomes via a metric μ, the object-level reasoning is already inherently comparative. The meta-level reasoning therefore focuses on the metric μ. The construction of this scheme follows the same procedure as in the previous cases; I summarise it below. The resulting meta-scheme for Consequentialism is presented in Box **??**.

- Creating a variable Arg to represent the object-level schemes from Consequentialism.
- Introducing a preference ordering between the cited metrics μ, which in natural language will be citing the justification for the preference.

Box 8 Meta-Scheme for Consequentialism

In the current circumstances R
Arg_1 advocates action A_1 based on μ_1
Arg_2 advocates action A_2 based on μ_2[a]
We should perform action A_1
Because there is a preference relation $\mu_2 \prec \mu_1$

[a] This can also capture cases of direct opposition, where the two competing arguments advocate for A and $\neg A$ respectively, with $A_2 = \neg A1$.

Critical Questions The CQs of this scheme are also constructed in a similar manner as before, with the appropriate adjustments. I will only elaborate the third CQ regarding a comparison of the evaluative metrics deployed in consequentialist schemes, as the rest are identical with the previous schemes.

The comparison of consequentialist metrics represents one of the most fundamental disagreements in moral philosophy, as different metrics can yield dramatically different conclusions about the same action. Consider the contrast between Classical Utilitarianism (which maximises total aggregate well-being) and Prioritarianism (which gives extra weight to the well-being of the worse-off). In policy decisions about healthcare resource allocation, Classical Utilitarianism might favour interventions that produce the greatest total health benefits across the population, while Prioritarianism would favour directing resources toward the most disadvantaged patients, even if this produces lower aggregate outcomes.

These metric preferences often reflect broader cultural and philosophical traditions. For instance, Average Utilitarianism (maximising average well-being per person) versus Total Utilitarianism (maximising total aggregate well-being) can lead to opposing conclusions about population policies. Western liberal democracies, with their emphasis on individual welfare, may be more inclined toward Average Utilitarianism when evaluating immigration policies, preferring to maintain higher average standards of living. In contrast, societies that emphasise collective flourishing might favour total utilitarianism, accepting some reduction in average welfare if it enables greater overall human flourishing through increased population. The critical question "Is the ordering $\mu_1 \succ \mu_2$ aligned with common practices?" thus probes whether the proposed metric preference reflects established evaluative norms within the relevant decision-making context, whether that be professional medical ethics, public policy frameworks, or broader cultural value systems.

The critical question for the *Meta-scheme for Consequentialism* are summarised as follows.

1. How well is the ordering $\mu_1 \succ \mu 2$ justified? What grounds justify this preference ordering?
2. Are A_1 and A_2 strictly incompatible, or can they be jointly or sequentially satisfied?
3. Is the ordering $\mu_1 \succ \mu 2$ aligned with common practices?
    - Are both metrics μ_1, μ_2 equally accepted?
4. Should this conflict be resolved by appealing to an alternative moral framework rather than this prioritisation?

8 Section Conclusion

This Section detailed the development of new argument schemes and critical questions specifically designed to enhance ethical reasoning. Through a two-stage approach combining empirical observations from Section **??** with philosophical analysis of normative ethical theories, I specified four object-level schemes corresponding to Virtue Ethics, Deontology, Consequentialism, and Rights-Based Ethics, along with comprehensive sets of associated critical questions. In addition, I developed a set of meta-schemes to model preference arguments reasoning over conflicts arising between arguments instantiating these new schemes.

The developed schemes offer a more nuanced and theoretically grounded representation of ethical reasoning than existing taxonomies. Each scheme captures the distinctive justificatory structure of its corresponding normative theory while maintaining formal rigour through *ASPIC*$^+$ representations, which specify precise attack relations between arguments through their critical questions. For instance, the *Action Scheme from Virtue Ethics* explicitly incorporates virtuous agents and moral virtues as variables, allowing for systematic examination of character-based justifications that existing schemes fail to capture. Similarly, the variations of the *Action Scheme from Deontology* reflect the nuanced distinctions within deontological reasoning that are crucial for accurate ethical modelling.

The schemes demonstrate clear improvements in modelling capacity compared to existing approaches. Most notably, the *Action Scheme from Consequentialism* explicitly separates empirical predictions about outcomes from normative evaluations of their desirability through the introduction of metric μ, thereby resolving the conflation problem present in Walton's *argument from consequences*. This separation allows for more precise challenges to consequentialist reasoning, distinguishing between disputes over factual predictions and disagreements over evaluative criteria. Example arguments from abortion debates, climate policy, and other ethically charged domains illustrate the practical applicability of the schemes and demonstrate how their associated critical questions can model sophisticated patterns of ethical reasoning that existing taxonomies cannot capture.

The comprehensive critical questions developed for each scheme serve multiple functions: eliciting supporting or attacking arguments, extending incomplete premises, and gathering contextual information. These questions draw from both empirical observations of real-world ethical debates and established philosophical critiques of the respective theories. For example, the critical questions for Virtue Ethics address key criticisms such as cultural relativism and the conflict problem, while those for Deontology and Rights-based Approach tackle issues such as universality and competing duties/rights. The consequentialist critical questions systematically address challenges including metric selection, consequence prediction, and temporal weighting, all central to evaluating outcome-based arguments.

Moreover, the development of these schemes revealed the need for meta-level reasoning in ethical discourse. Many critical questions point to conflicts within a normative framework (e.g., conflicting duties within the deontological framework) and between competing normative frameworks. Such conflicts cannot

be resolved at the object level; they require explicit reasoning about ethical considerations should take precedence in particular circumstances. This Section addressed the first of these needs by developing meta-schemes specialised for intra-theoretical comparison, enabling structured reasoning over preferences between competing rules, rights, virtues, or consequences within the corresponding framework.

Overall, the schemes developed in this Section represent a significant advance in computational Ethics and argumentative reasoning. They provide the foundational infrastructure for more sophisticated ethical deliberation systems while maintaining the theoretical rigour necessary for principled moral reasoning. At the same time, the analysis points toward the next stage of development: meta-schemes that enable structured reasoning over preferences between competing normative theories (e.g., when consequentialist arguments clash with deontological ones). Section **??** outlines directions for future research, in which the foundations presented here can be extended to systematically develop such schemes for inter-theoretical comparisons, along with other forms of meta-reasoning commonly observed in ethical deliberation.

References

[1] Robert Merrihew Adams. *A theory of virtue: Excellence in being for the good*. Clarendon Press, 2008.

[2] Larry Alexander and Michael Moore. Deontological Ethics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021.

[3] Thomas Aquinas. *Summa Theologica, Second Part of the Second Part (II-II)*. Benziger Bros., New York, 1947. Translated by Fathers of the English Dominican Province.

[4] Aristotle. *Nicomachean Ethics*. Oxford University Press, Oxford, 2004.

[5] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Computational representation of practical argument. *Synthese*, 152:157–206, 2006.

[6] Katie Atkinson and Trevor JM Bench-Capon. Value-based argumentation. *FLAP*, 8(6):1543–1588, 2021.

[7] Trevor Bench-Capon and Katie Atkinson. Abstract argumentation and values. *Argumentation in artificial intelligence*, pages 45–64, 2009.

[8] Trevor J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[9] Stephen F. Bigger. The family laws of leviticus 18 in their setting. *Journal of Biblical Literature*, 98(2):187–203, 1979.

[10] Nicholas J Campbell. A comparative interpretation of the old testament prohibited mixtures: Mixed breeding in leviticus 19: 19. *The Catholic Biblical Quarterly*, 85(2):199–218, 2023.

[11] Roger Crisp. Well-Being. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2025 edition, 2025.

[12] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.

[13] David Heyd. Supererogation. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2024 edition, 2024.

[14] Brad Hooker. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford University Press, Oxford, 2000.

[15] Rosalind Hursthouse and Glen Pettigrove. Virtue Ethics. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2022 edition, 2022.

[16] Jillian J Jordan, Roseanna Sommers, Paul Bloom, and David G Rand. Why do we hate hypocrites? evidence for a theory of false signaling. *Psychological science*, 28(3):356–368, 2017.

[17] Immanuel Kant. On a supposed right to tell a lie from altruistic motives. In Mary J. Gregor, editor, *Practical Philosophy*, pages 605–615. Cambridge University Press, Cambridge, 1996. Originally published in 1797.

[18] Nadin Kökciyan, Isabel Sassoon, Elizabeth Sklar, Sanjay Modgil, and Simon Parsons. Applying metalevel argumentation frameworks to support medical decision making. *IEEE Intelligent Systems*, 36(2):64–71, 2021.

[19] James Lenman. Consequentialism and cluelessness. *Philosophy & public affairs*, 29(4):342–370, 2000.

[20] Robert B. Louden. On some vices of virtue ethics. In Roger Crisp and Michael Slote, editors, *Virtue Ethics*, pages 201–216. Oxford University Press, Oxford, 1997.

[21] William MacAskill. *What We Owe The Future: The Sunday Times Bestseller*. Simon and Schuster, 2022.

[22] Alasdair MacIntyre. *After Virtue*. Duckworth, London, 2 edition, 1985.

[23] Alasdair MacIntyre. *After virtue*. A&C Black, 2013.

[24] Tim Mulgan. *The Demands of Consequentialism*. Oxford University Press, Oxford, 2001. Online edition, Oxford Academic, 31 Oct 2023.

[25] Richard J Norman. The moral philosophers: An introduction to ethics. 1998.

[26] Derek Parfit. *Reasons and persons*. Oxford University Press, 1987.

[27] Isabel Sassoon, Nadin Kökciyan, Sanjay Modgil, and Simon Parsons. Argumentation schemes for clinical decision support. *Argument & Computation*, 12(3):329–355, 2021.

[28] Henry Shue. Torture. *Philosophy & Public Affairs*, pages 124–143, 1978.

[29] Sophocles. *Antigone*. Prestwick House Inc, 2005.

[30] Judith Jarvis Thomson. The trolley problem. *The Yale Law Journal*, 94(6):1395–1415, 1985.

[31] Judith Jarvis Thomson. *Goodness and advice*. Princeton University Press, 2003.

[32] Pancho Tolchinsky, Ulises Cortés, Sanjay Modgil, Francisco Caballero, and Antonio López-Navidad. Increasing human-organ transplant availability: Argumentation-based agent deliberation. *IEEE Intelligent Systems*, 21(6):30–37, 2006.

[33] Pancho Tolchinsky, Sanjay Modgil, Katie Atkinson, Peter McBurney, and Ulises Cortés. Deliberation dialogues for reasoning about safety critical actions. *Autonomous Agents and Multi-Agent Systems*, 25(2):209–259, 2012.

[34] Liezl Van Zyl. Virtue ethics and the trolley problem. *The Trolley Problem*, pages 116–134, 2023.

[35] Wendi Zhou, Ameer Saadat-Yazdi, and Nadin Kökciyan. Aspect-based opinion summarization with argumentation schemes. In Elena Chistova, Philipp Cimiano, Shohreh Haddadan, Gabriella Lapesa, and Ramon Ruiz-Dolz, editors, *Proceedings of the 12th Argument mining Workshop*, pages 116–125, Vienna, Austria, July 2025. Association for Computational Linguistics.