Auton Agent Multi-Agent Syst
DOI 10.1007/s10458-007-9016-7

Prioritized conditional imperatives: problems and a new proposal

Jörg Hansen

Abstract The sentences of deontic logic may be understood as describing what an agent ought to do when faced with a given set of norms. If these norms come into conflict, the best the agent can be expected to do is to follow a maximal subset of the norms. Intuitively, a priority ordering of the norms can be helpful in determining the relevant sets and resolve conflicts, but a formal resolution mechanism has been difficult to provide. In particular, reasoning about prioritized conditional imperatives is overshadowed by problems such as the 'order puzzle' that are not satisfactorily resolved by existing approaches. The paper provides a new proposal as to how these problems may be overcome.

Keywords Deontic logic · Default logic · Priorities · Logic of imperatives

1 Drinking and driving

Imagine you have been invited to a party. Before the event, you become the recipient of various imperative sentences:

(1) Your mother says: if you drink anything, then don't drive.
(2) Your best friend says: if you go to the party, then you do the driving.
(3) Some acquaintance says: if you go to the party, then have a drink with me.

Suppose that as a rule you do what your mother tells you—after all, she is the most important person in your life. Also, the last time you went to a party your best friend did the driving, so it really is your turn now. You can enjoy yourself without a drink, though it would be nice to have a drink with your acquaintance—your best friend would not mind if you had one drink, and your acquaintance does not care that you may be driving—but your mother would not approve of such a behavior. Making up your mind,

J. Hansen (✉)
Institut für Philosophie, University of Leipzig, Beethovenstraße 15, 04107 Leipzig, Germany
e-mail: jhansen@uni-leipzig.de

(4) You go to the party.

I think that intuitively it is quite clear what you must do: obey your mother and your best friend, and hence do the driving and not accept your acquaintance's invitation. However, it is not so clear what formal algorithm could explain this reasoning.

An example of a similar form was first employed in epistemic logic,[1] and has been termed the 'order puzzle' (cf. Horty [21]). For the epistemic version, consider the following sentences:

(5) You remember from physics: if you are in a car, lightning won't strike you.
(6) The coroner tells you: he was struck by lightning.
(7) Your neighbor says: he must have been drinking and driving.

Suppose that driving includes being in a car, that you firmly believe in what you remember from physics, that you believe that information by medical officers is normally based on competent investigation, and that you usually don't question your neighbor's observations, but think that sometimes she is just speculating. It seems quite clear what happens: you keep believing what you remember from school, and don't doubt what the coroner told you, but question your neighbor's information, maybe answering: "This can't be true, as the authorities found he was struck by lightning, and you can't be struck by lightning in a car."

In both cases, the problem as to how the underlying reasoning can be formally reconstructed seems so far unsolved. Both involve a priority ordering of the sentences involved. While the paper discusses the imperative side of things from the angle of philosophical (deontic) logic, its solution seems also relevant for the similarly structured problems of conditional beliefs and desires and the modeling of agent reasoning in the face of such conditionals. Sections 2 and 3 present the formal framework for the discussion of conditional imperatives and resulting obligations. Section 4 examines various proposals as to how a priority ordering may be used to resolve conflicts, it turns out that all of these do not solve the 'order puzzle.' A postulate at the beginning of Sect. 5 summarizes our intuitions in this matter, but also delegates the solution to the problem to a proper definition of what it means that conditional imperatives conflict in a given situation. Three such definitions are studied, of which the third seems to solve the problems. Section 6 gives theorems of a corresponding deontic logic and Sect. 7 points at remaining problems for the representation of conditional imperatives. Section 8 concludes.

2 Formal preliminaries

To formally discuss problems such as the one presented above, I shall use a simple framework: let I be a set of objects, they are meant to be (conditional) imperatives. Two functions g and f associate with each imperative an antecedent and a consequent—these are sentences from the language of a basic logic that here will be the language $\mathscr{L}_{PL}$ of propositional logic.[2] $g(i)$ may be thought of as describing the 'grounds', or circumstances in which the consequent of i is to hold, and $f(i)$ as associating the sentence that describes what must be

[1] Cf. Rintanen [34] p. 234, who in turn credits Gerhard Brewka with its invention.

[2] PL is based on a language $\mathscr{L}_{PL}$, defined from a set of proposition letters $Prop = \{p_1, p_2, ...\}$, Boolean connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ and brackets (,) as usual. The truth of a $\mathscr{L}_{PL}$-sentence (I use upper case letters $A, B, C, ...$) is defined recursively using valuations $v : Prop \rightarrow \{1, 0\}$ (I write $v \models A$), starting with $v \models p$ iff $v(p) = 1$ and continuing as usual. If $A \in \mathscr{L}_{PL}$ is true for all valuations it is called a tautology. PL is the set of all tautologies, and used to define provability, consistency and derivability (I write $\Gamma \vdash_{PL} A$) as usual. $\top$ is an arbitrary tautology, and $\bot$ is $\neg\top$.

the case if the imperative i is satisfied, its 'deontic focus' or 'demand.'[3] In accordance with tradition (cf. Hofstadter and McKinsey [19]), I write $A \Rightarrow !B$ for an $i \in I$ with $g(i) = A$ and $f(i) = B$, and $!A$ means an unconditional imperative $\top \Rightarrow !A$. Note that $A \Rightarrow !B$ is just the name for a conditional imperative that demands B to be made true in a situation where A is true—it is not an object that is assigned truth values. A useful construction is the 'materialization' $m(i)$ of an imperative i, which is the material implication $g(i) \rightarrow f(i)$ that may be thought of as corresponding to a conditional imperative. For any $i \in I$ and $\Delta \subseteq I$, instead of $f(i), g(i), m(i), f(\Delta), g(\Delta)$ and $m(\Delta)$, I may use the superscripted i^f, i^g, i^m, Δ^f, Δ^g and Δ^m for better readability.

Let $\mathcal{I}$ be a tuple $\langle I, f, g\rangle$, let $W \subseteq \mathscr{L}_{PL}$ be a set of sentences, representing 'real world facts,' and $\Delta \subseteq I$ be a subset of the imperatives: then we define

$$\mathit{Triggered}_{\mathcal{I}}(W, \Delta) = \{i \in \Delta | W \vdash_{PL} g(i)\}.$$

So an imperative $i \in \Delta$ is triggered if its antecedent is true given W. Tradition wants it that a conditional imperative can only be fulfilled or violated if its condition is the case.[4] So I define:

$$\mathit{Satisfied}_{\mathcal{I}}(W, \Delta) = \{i \in \Delta | W \vdash_{PL} i^g \wedge i^f\},$$
$$\mathit{Violated}_{\mathcal{I}}(W, \Delta) = \{i \in \Delta | W \vdash_{PL} i^g \wedge \neg i^f\},$$

An imperative in $\mathit{Satisfied}_{\mathcal{I}}(W, \Delta)$ [$\mathit{Violated}_{\mathcal{I}}(W, \Delta)$] is called satisfied [violated] given the facts W. It is of course possible that an imperative is neither satisfied nor violated given the facts W. If an imperative is triggered, but not violated, we call the imperative satisfiable:

$$\mathit{Satisfiable}_{\mathcal{I}}(W, \Delta) = \{i \in \mathit{Triggered}_{\mathcal{I}}(W, \Delta) | W \nvdash_{PL} \neg i^f\}.$$

Moreover, the following definition will play a major rôle in what follows:

$$\mathit{Obeyable}_{\mathcal{I}}(W, \Delta) = \{\Gamma \subseteq \Delta | \Gamma^m \cup W \nvdash_{PL} \bot\}.$$

So a subset Γ of Δ is obeyable given W iff it is not the case that for some $\{i_1, \ldots, i_n\} \subseteq \Gamma$ we have $W \vdash_{PL} (i_1^g \wedge \neg i_1^f) \vee \ldots \vee (i_n^g \wedge \neg i_n^f)$: otherwise we know that whatever we do, i.e. given any maxiconsistent subset V of $\mathscr{L}_{PL}$ that extends $W \subseteq V$, at least one imperative in Γ is violated.[5] We speak of a *conflict of imperatives* when the triggered imperatives cannot all be satisfied given the facts W, i.e. when $\mathit{Triggered}_{\mathcal{I}}(W, \Delta)^f \cup W \vdash_{PL} \bot$. More generally speaking I will also call imperatives conflicting if they are not obeyable in the given situation.

As prioritized conditional imperatives are our concern here, we let all imperatives in I be ordered by some priority relation $< \subseteq I \times I$. The relation $<$ is assumed to be a strict partial order on I, i.e. $<$ is irreflexive and transitive, and additionally we assume $<$ to be well-founded, i.e. infinite descending chains are excluded. For any $i_1, i_2 \in I, i_1 < i_2$ means that i_1 takes priority over i_2 (ranks higher than i_2, is more important than i_2, etc.). A tuple $\langle I, f, g\rangle$ will be called a *conditional imperative structure*, and $\langle I, f, g, <\rangle$ a *prioritized conditional imperative structure*. If all imperatives in I are unconditional, we may drop any reference to the relation g in the tuples and call these *basic imperative structures* and *prioritized imperative structures* respectively.

[3] In analogy to Reiter's default logic one might add a third function e that describes exceptional circumstances in which the imperative is not to be applied. I will not address this additional complexity here.

[4] Cf. Rescher [33], Sosa [38], van Fraassen [9]. Also cf. Greenspan [11]: "Oughts do not arise, it seems, until it is too late to keep their conditions from being fulfilled."

[5] Terms differ here, e.g. Downing [8] uses the term 'compliable' instead of 'obeyable.'

3 Deontic concepts

Given a set of imperatives, one may truly or falsely state that their addressee must, or must not, perform some act or achieve some state of affairs according to what the addressee was ordered to do. Regarding the 'drinking and driving' example, I think it is true that the agent ought to do the driving, as this is what the second-ranking imperative, uttered by the agent's best friend, requires her to do, but that it would be false to say that the agent ought to drink and drive. Statements that something ought to be done or achieved are called 'normative' or 'deontic statements,' and the ultimate goal of deontic logic is to find a logical semantics that models the situation and defines the deontic concepts in such a way that the formal results coincide with our natural inclinations in the matter.

3.1 Deontic operators for unconditional imperatives

For unconditional imperatives, such definitions are straightforward. Given a basic imperative structure $\mathcal{I} = \langle I, f \rangle$, a monadic deontic O-operator, that formalizes 'it ought to be that A (is realized)' by OA, is defined by

$$(td\text{-}m1)\quad \mathcal{I} \models OA \text{ if and only if (iff) } I^f \vdash_{PL} A.$$

So obligation is defined in terms of what the satisfaction of all imperatives logically implies.[6] With the usual truth definitions for Boolean operators, it can easily be seen that such a definition produces a normal modal operator, i.e. one that is defined by the following axiom schemes plus *modus ponens*:

$$\begin{array}{ll} \text{(Ext)} & \text{If } \vdash_{PL} A \leftrightarrow B, \text{ then } OA \leftrightarrow OB \text{ is a theorem.} \\ \text{(M)} & O(A \wedge B) \to (OA \wedge OB) \\ \text{(C)} & (OA \wedge OB) \to O(A \wedge B) \\ \text{(N)} & O\top \end{array}$$

Furthermore, (*td-m*1) defines standard deontic logic *SDL*, which adds

$$\text{(D)}\quad OA \to \neg O \neg A$$

iff the imperatives are assumed to be non-conflicting and so I^f is *PL*-consistent, i.e. $I^f \nvdash_{PL} \bot$. It is immediate that in the case of conflicts, (*td-m*1) pronounces everything as obligatory, and in particular defines $O\bot$ true, thus making the impossible obligatory. If conflicts are not excluded, a solution is to only consider (maximal) subsets of the imperatives whose demands are consistent and define the O-operator with respect to these (I write $I \curlywedge C$ for the set of all 'C-remainders,' i.e. maximal subsets Γ of I such that $\Gamma^f \nvdash_{PL} C$):

$$(td\text{-}m2)\ \mathcal{I} \models OA \text{ iff } \forall \Gamma \in I \curlywedge \bot : \Gamma^f \vdash_{PL} A$$

Quite similarly, a dyadic deontic operator $O(A/C)$, meaning that A ought to be true given that C is true, can be defined with respect to the maximal subsets of imperatives that do not conflict in these circumstances:

$$(td\text{-}d1)\ \mathcal{I} \models O(A/C) \text{ iff } \forall \Gamma \in I \curlywedge \neg C : \Gamma^f \vdash_{PL} A$$

[6] Such a definition of obligation was proposed e.g. by Alchourrón and Bulygin [1].

So A is obligatory given that C is true if A is what the imperatives in any $\neg C$-remainder demand.[7] In the case of conflicts, this definition produces a "disjunctive solution:" e.g. if there are two imperatives $!A$ and $!B$ with $\vdash_{PL} C \rightarrow (A \rightarrow \neg B)$, then neither $O(A/C)$ nor $O(B/C)$ but $O(A \vee B/C)$ is true.[8]

Often, we want to use the information that we have about the circumstances also for reasoning about the obligations in these circumstances. E.g. if the set of imperatives is $\{!(p_1 \vee p_2)\}$, ordering me to either send you a card or phone you, and I cannot send you a card, i.e. $\neg p_1$ is true, I should be able to conclude that I should phone you, and so $O(p_2/\neg p_1)$ should be true. Such 'circumstantial reasoning' is achieved by the following change to the truth definition:

$$(\textit{td-d}1^+)\ \mathcal{I} \models O(A/C) \text{ iff } \forall\Gamma \in I \curlywedge \neg C : \Gamma^f \cup \{C\} \vdash_{PL} A$$

So A is obligatory given C is (invariably) true iff all maximal subsets of the imperatives' demands (the imperatives' associated descriptive sentences) that are consistent with the circumstances C, plus C, derive A. With the usual truth conditions for Boolean operators, a semantics that employs (*td-d*1^+) has a sound and (weakly) complete axiom system *PD* that extends the system *P* of Kraus et al. [22], defined by these axiom schemes

(DExt)	If $\vdash_{PL} A \leftrightarrow B$ then $O(A/C) \leftrightarrow O(B/C)$ is a theorem.
(DM)	$O(A \wedge B/C) \rightarrow (O(A/C) \wedge O(B/C))$
(DC)	$O(A/C) \wedge O(B/C) \rightarrow O(A \wedge B/C)$
(DN)	$O(\top/C)$
(ExtC)	If $\vdash_{PL} C \leftrightarrow D$ then $O(A/C) \leftrightarrow O(A/D)$ is a theorem.
(CCMon)	$O(A \wedge D/C) \rightarrow O(A/C \wedge D)$
(CExt)	If $\vdash_{PL} C \rightarrow (A \leftrightarrow B)$ then $O(A/C) \leftrightarrow O(B/C)$ is a theorem.
(Or)	$O(A/C) \wedge O(A/D) \rightarrow O(A/C \vee D)$

with the additional (restricted, dyadic) 'deontic' axiom scheme

$$\text{(DD-R) If } \nvdash_{PL} \neg C \text{ then } \vdash_{PD} O(A/C) \rightarrow \neg O(\neg A/C)$$

(sometimes called "preservation of classical consistency"), hence the name *PD*.[9]

3.2 Deontic operators for conditional imperatives

Unlike their unconditional counterparts, conditional imperatives have been found hard to reason about. von Wright [41] called conditional norms the "touchstone of normative logic," and van Fraassen [9] wrote with regard to logics for conditional imperatives: "There may

[7] Though statements like $O(\neg C/C)$ are syntactically well-formed, they are thus defined false for any possible situation C—this is the same for any dyadic deontic logic since Hansson [15] and Lewis [23] (for the motivation cf. [26] pp. 158–159). Similarly, imperatives like $C \Rightarrow !\neg C$ are treated as violated as soon as they are triggered by the facts. There exist meaningful natural-language imperatives like 'close the window if it is open,' but I think that in these the proposition in the antecedent is different from the negation of the one corresponding to the consequent, in that the second refers to a different point of time ('see to it that the window is closed some time in the near future if it is open now'), so they should not be represented by $C \Rightarrow !\neg C$.

[8] For alternative solutions to the problem of conflicts cf. Goble [10] and Hansen [12], [13].

[9] For proofs, and an additional "credulous ought" that defines $O(A/C)$ true if the truth of A is required to satisfy all imperatives in *some* $\neg C$-remainder, cf. Hansen [13].

be systematic relations governing this moral dynamics, but I can only profess ignorance of them."

Representing a conditional imperative as an unconditional imperative that demands a material conditional to be made true yields undesired results. Most notorious is the problem of contraposition: consider a set I with the only imperative $!(p_1 \rightarrow p_2)$, meaning e.g. 'if it rains, take an umbrella.' (*td-d*1) makes true $O(p_2/p_1)$, but also $O(\neg p_1/\neg p_2)$, so if you cannot take your umbrella (your wife took it) you must see to it that it does not rain, which is hardly what the speaker meant you to do. One may think that such problems arise from the fact that antecedents of conditional imperatives often describe states of the affairs that the agent is not supposed to, and often cannot, control. But consider the set $\{!(p_1 \rightarrow p_2), !(\neg p_1 \rightarrow p_3)\}$, it yields $O(p_2/\neg p_3)$ with (*td-d*1). Here, p_2 is what the consequent of some imperative demands, so it supposedly describes something the agent can control. Now let the imperatives be interpreted as ordering me to wear a rain coat if it rains, and my best suit if it does not: it is clear nonsense that I am obliged to wear a raincoat given that I can't wear my best suit (e.g. it is in the laundry). Such problems are the reason why we use special models for conditional imperatives that separate antecedents and consequents (conditional imperative structures), and write $p_1 \Rightarrow !p_2$ instead of $!(p_1 \rightarrow p_2)$. But this only delegates the problem from the level of representation to that of semantics, where now new truth definitions must be found.

Let $\mathcal{I} = \langle I, f, g \rangle$ be a conditional imperative structure, and let us ignore for the moment the further complication of possible conflicts between imperatives. Then the following seems a natural way to define what ought to be the case in circumstances where C is assumed to be true:

$$(\textit{td-cd}1)\ \mathcal{I} \models O(A/C) \text{ iff } [\textit{Triggered}_{\mathcal{I}}(\{C\}, I)]^f \vdash_{PL} A$$

So dyadic obligation is defined in terms what is necessary to satisfy all imperatives that are triggered in the assumed circumstances. E.g. if $I = \{p_1 \Rightarrow !p_2\}$, with its only imperative interpreted as "if you have a cold, stay in bed," then $O(p_2/p_1)$ truly states that I must stay in bed given that I have a cold.

Like in the unconditional case, it seems important to be able to use 'circumstantial reasoning,' i.e. employ the information about the situation not only to determine if an imperative is triggered, but also for reasoning with its consequent. E.g. if the set of imperatives is $\{p_1 \Rightarrow !(p_2 \vee p_3)\}$, with its imperative interpreted as expressing "if you have a cold, either stay in bed or wear a scarf," one would like to obtain $O(p_3/p_1 \wedge \neg p_2)$, expressing that given that I have a cold and don't stay in bed, I must wear a scarf. So (*td-cd*1) may be changed into

$$(\textit{td-cd}1^+)\ \mathcal{I} \models O(A/C) \text{ iff } [\textit{Triggered}_{\mathcal{I}}(\{C\}, I)]^f \cup \{C\} \vdash_{PL} A.$$

Though the step from (*td-cd*1) to (*td-cd*1$^+$) seems quite reasonable, such definitions have also been criticized for defining the assumed circumstances as obligatory. In the above example, (*td-cd*1$^+$) also makes true $O(p_1/p_1 \wedge \neg p_2)$, so given that you have a cold it is true that you ought to have it. The criticism loses much of its edge in the present setting, where one can point to the distinction between imperatives (there is no imperative that demands p_1) and ought sentences that describe what must be true given the facts and the satisfaction of all triggered imperatives: then the truth of $O(p_1/p_1)$ seems no more paradoxical than the truth of $O\top$ that is accepted in most systems of deontic logic.

3.3 Further modifications

In Makinson and van der Torre's [25] more general theory of 'input/output logic,' (*td-cd*1) is termed 'simple-minded output,' and (*td-cd*1^+) is its 'throughput version.'[10] As the names suggests, the authors also discuss more refined operations, which again might be considered for reasoning about conditional imperatives. One modification addresses the possibility of 'reasoning by cases' that e.g. makes true $O(p_2 \vee p_4/p_1 \vee p_3)$ for a set of imperatives $I = \{p_1 \Rightarrow !p_2, p_3 \Rightarrow !p_4\}$. This may be achieved by the following definition, where $\mathscr{L}_{PL} \bot \neg C$ is the set of all maximal subsets of the language $\mathscr{L}_{PL}$ that are consistent with C:[11]

$$(\textit{td-cd}2)\ \mathcal{I} \models O(A/C) \text{ iff } \forall V \in \mathscr{L}_{PL} \bot \neg C : [\textit{Triggered}_{\mathcal{I}}(V, I)]^f \vdash_{PL} A$$

In the example, each set $V \subset \mathscr{L}_{PL}$ that is maximally consistent with $p_1 \vee p_3$ either contains p_1, then $p_1 \Rightarrow !p_2$ is triggered and so p_2 and also $p_2 \vee p_4$ is implied by $[\textit{Triggered}_{\mathcal{I}}(V, I)]^f$, or it contains $\neg p_1$, but then it cannot also contain $\neg p_3$ and so must contain p_3, so $p_3 \Rightarrow !p_4$ is triggered and therefore p_4 and also $p_2 \vee p_4$ implied, so for all sets V, $p_2 \vee p_4$ is implied and so $O(p_2 \vee p_4/p_1 \vee p_3)$ made true.

In order to add 'circumstantial reasoning' to (*td-cd*3)—or, in Makinson and van der Torre's terms, for its 'throughput version'—one might, in the vein of (*td-d*1^+) and (*td-cd*1^+), try this definition:

$$(\textit{td-cd}2^-)\ \mathcal{I} \models O(A/C) \text{ iff } \forall V \in \mathscr{L}_{PL} \bot \neg C : [\textit{Triggered}_{\mathcal{I}}(V, I)]^f \cup \{C\} \vdash_{PL} A$$

But the definition seems too weak. Consider the set $\{p_1 \Rightarrow !(\neg p_2 \vee p_4), p_3 \Rightarrow !p_4\}$ and the situation $(p_1 \wedge p_2) \vee p_3$. We would expect a reasoning as follows: in this situation, either $p_1 \wedge p_2$ is true, so the first imperative is triggered but we cannot satisfy it by bringing about $\neg p_2$, and so must bring about p_4. Or p_3 is true, then the second imperative is triggered and we must again bring about p_4. So we must bring about p_4 in the given situation. But the definition fails to make true $O(p_4/(p_1 \wedge p_2) \vee p_3)$. Like Makinson and van der Torre [25], I therefore combine reasoning by cases with a stronger version of throughput:

$$(\textit{td-cd}2^+)\ \mathcal{I} \models O(A/C) \text{ iff } \forall V \in \mathscr{L}_{PL} \bot \neg C : [\textit{Triggered}_{\mathcal{I}}(V, I)]^f \cup V \vdash_{PL} A$$

As is easy to see, this resolves the difficulty: for $\{p_1 \Rightarrow !(\neg p_2 \vee p_4), p_3 \Rightarrow !p_4\}$, $O(p_4/(p_1 \wedge p_2) \vee p_3)$ is now true, as desired. But this modification has an unwanted consequence: it makes reasoning about conditional imperatives collapse into reasoning about consequences of their materializations (cf. [26] p. 156):

Observation 1 *By* $(td - cd2^+)$, $\mathcal{I} \models O(A/C)$ *iff* $m(I) \cup \{C\} \vdash_{PL} A$.

Proof For the right-to-left direction, for any imperative $i \in I$ and any set $V \in \mathscr{L}_{PL} \bot \neg C$, either V includes $g(i)$, so $i \in \textit{Triggered}_{\mathcal{I}}(V, I)$ and therefore $[\textit{Triggered}_{\mathcal{I}}(V, I)]^f \vdash_{PL} g(i) \rightarrow f(i)$, or it does not include $g(i)$, but then it includes $\neg g(i)$ by maximality, hence $V \vdash_{PL} g(i) \rightarrow f(i)$. So $[\textit{Triggered}_{\mathcal{I}}(V, I)]^f \cup V \vdash_{PL} g(i) \rightarrow f(i)$. For the left-to-right direction, if $m(I) \cup \{C\} \nvdash_{PL} A$ then $m(I) \cup \{C\} \cup \{\neg A\}$ is consistent, so there is a $V \in \mathscr{L}_{PL} \bot \neg C$ such that $m(I) \cup \{C\} \cup \{\neg A\} \subseteq V$. It is immediate that for each $i \in$

[10] If I resembles the generating set G of input/output logic, then $O(A/C)$ means that A is an output given the input C (Makinson and van der Torre write $A \in out(G, \{C\})$). Though these authors liken their generating set G to a body of conditional norms, it should be noted that they do not themselves introduce dyadic deontic operators.

[11] Makinson and van der Torre's [25] call the resulting operator 'basic output,' of which a syntactical version was first presented by Świrydowicz [39] p. 32.

$Triggered_{\mathcal{I}}(V, I)$, $m(I) \cup V \vdash_{PL} f(i)$, so if $[Triggered_{\mathcal{I}}(V, I)]^f \cup V \vdash_{PL} A$ then $m(I) \cup V \vdash_{PL} A$ and since $m(I) \subseteq V$ also $V \vdash_{PL} A$. Since V was consistent and included $\neg A$, it cannot also derive A, and so by contraposition $[Triggered_{\mathcal{I}}(V, I)]^f \cup V \nvdash_{PL} A$.

But such an equivalence makes all the problems of identifying conditional imperatives with unconditional ones that demand their materializations reappear, in particular the problem of contraposition.[12] So it seems we must choose between 'reasoning by cases' and 'circumstantial reasoning.' Another modification that these authors consider is that of 'reusable output:' when an imperative is triggered that demands A, and A is the trigger for some imperative $A \Rightarrow !B$, then we also ought to do B. Such a modification can easily be incorporated into a truth definition and its 'throughput' version:

$$(td\text{-}cd3)\quad \mathcal{I} \models O(A/C) \text{ iff } [Triggered^*_{\mathcal{I}}(\{C\}, I)]^f \vdash_{PL} A$$

$$(td\text{-}cd3^+)\quad \mathcal{I} \models O(A/C) \text{ iff } [Triggered^*_{\mathcal{I}}(\{C\}, I)]^f \cup \{C\} \vdash_{PL} A$$

where $Triggered^*_{\mathcal{I}}(W, \Gamma)$ means the smallest subset of $\Gamma \subseteq I$ such that for all $i \in \Gamma$, if $[Triggered^*_{\mathcal{I}}(W, \Gamma)]^f \cup W \vdash_{PL} g(i)$ then $i \in Triggered^*_{\mathcal{I}}(W, \Gamma)$. Moreover, the two modifications of 'reasoning by cases' and 'reusable output' can be combined to produce the following definition and its 'throughput' variant:

$$(td\text{-}cd4)\quad \mathcal{I} \models O(A/C) \text{ iff } \forall V \in \mathscr{L}_{PL}\bot\neg C : [Triggered^*_{\mathcal{I}}(V, I)]^f \vdash_{PL} A$$

$$(td\text{-}cd4^+)\quad \mathcal{I} \models O(A/C) \text{ iff } \forall V \in \mathscr{L}_{PL}\bot\neg C : [Triggered^*_{\mathcal{I}}(V, I)]^f \cup V \vdash_{PL} A$$

The topic of 'reusable output' is discussed under the name of 'deontic detachment' in the deontic logic literature, and there is no agreement whether such a procedure is admissible (Makinson [24] p. 43 argues in favor, whereas Hansson [16] p. 155 disagrees). E.g. let $I = \{!p_1, p_1 \Rightarrow !p_2\}$, and for its interpretation assume that it is imperative for the proper execution of your job that you develop novel methods, which make you eligible for a bonus, and that if you develop such novel methods you owe it to yourself to apply for the bonus. Truth definitions that accept 'deontic detachment' make true $O(p_2/\top)$, and so tell us that you ought to apply for the bonus, which seems weird since it may be that you never invent anything. However, proponents of deontic detachment may argue that in such a situation, $O(p_1 \wedge p_2/\top)$ should hold, i.e. you ought to invent new methods *and* apply for the bonus, and that the reluctance to also accept $O(p_2/\top)$ is—like the inference from "you ought to put on your parachute and jump" to "you ought to jump"—just a variant of Ross' Paradox that is usually considered harmless.

For (*td-cd*4) we once again obtain $O(p_2/\neg p_3)$ for $I = \{p_1 \Rightarrow !p_2, \neg p_1 \Rightarrow !p_3\}$: for any $V \in \mathscr{L}_{PL}\bot p_3$, $\neg p_3 \in V$, furthermore either $p_1 \in V$ and so $p_1 \Rightarrow !p_2 \in Triggered^*_{\mathcal{I}}(V, I)$, or $\neg p_1 \in V$, then $\neg p_1 \Rightarrow !p_3 \in Triggered^*_{\mathcal{I}}(V, I)$, and since $\{p_3\} \cup \{\neg p_3\} \vdash_{PL} p_1$, again $p_1 \Rightarrow !p_2$ is in $Triggered^*_{\mathcal{I}}(V, I)$, hence $[Triggered^*_{\mathcal{I}}(V, I)]^f \vdash_{PL} p_2$ for all $V \in \mathscr{L}_{PL}\bot p_3$. But as we saw above, this result seems counterintuitive.[13] Note that (*td-cd*4$^+$) is again equivalent to $\mathcal{I} \models O(A/C)$ iff $m(I) \cup \{C\} \vdash_{PL} A$ and thus to (*td-cd*2$^+$) (cf. [25] observation 16; [26], p. 156). □

Observation 2 *By* (*td-cd*4$^+$), $\mathcal{I} \models O(A/C)$ *iff* $m(I) \cup \{C\} \vdash_{PL} A$.

Proof Similar to the proof of observation 1. For the left-to-right direction, use that for each $i \in Triggered^*_{\mathcal{I}}(V, I)$, $m(I) \cup V \vdash_{PL} f(i)$, which is immediate. □

[12] (*td-cd*2$^-$) does not fare much better: though it does not include contraposition, it again makes $O(p_2/\neg p_3)$ true for $I = \{p_1 \Rightarrow !p_2, \neg p_1 \Rightarrow !p_3\}$, which is counterintuitive.

[13] With respect to their out_4-operation that corresponds to (*td-cd*4), Makinson and van der Torre [25] speak of a 'ghostly contraposition'.

3.4 Operators for prioritized conditional imperatives

This paper focuses on prioritized conditional imperatives, and for these there is a further hurdle to finding the proper truth definitions for deontic concepts. Priorities are only required if the imperatives cannot all be obeyed—otherwise there is no reason not to obey all, and the priority ordering is not used. So the truth definitions must be able to deliver meaningful results for possibly conflicting imperatives. The intuitive idea is to use the information in the ordering to choose subsets of the set of imperatives under consideration that contain only the more important imperatives and leave out less important, conflicting ones, so that the resulting 'preferred subset' (or rather, subsets, since the choice may not always be uniquely determined by the ordering) only contains imperatives that do not conflict in the given situation. More generally, let $\mathcal{I}$ be a prioritized conditional imperative structure $\langle I, g, f, <\rangle$, and let Δ be a subset of I. Then $\mathscr{P}_{\mathcal{I}}(W, \Delta)$ contains just the subsets of Δ that are thus preferred given the world facts W. The above truth definitions can then be adapted such that they now describe something as obligatory iff it is so with respect to all the preferred subsets of the imperatives, i.e. they take on the following forms:

$$\mathcal{I} \models O(A/C) \text{ iff } \forall\Gamma \in P_{\mathcal{I}}(\{C\}, I):$$

$$\begin{array}{lr}
(\textit{td-pcd}1) & [\textit{Triggered}_{\mathcal{I}}(\{C\}, \Gamma)]^f \vdash_{PL} A, \\
(\textit{td-pcd}1^+) & [\textit{Triggered}_{\mathcal{I}}(\{C\}, \Gamma)]^f \cup \{C\} \vdash_{PL} A, \\
(\textit{td-pcd}2) & \forall V \in \mathscr{L}_{PL}\bot\neg C : [\textit{Triggered}_{\mathcal{I}}(V, \Gamma)]^f \vdash_{PL} A, \\
(\textit{td-pcd}2^+) & \forall V \in \mathscr{L}_{PL}\bot\neg C : [\textit{Triggered}_{\mathcal{I}}(V, \Gamma)]^f \cup V \vdash_{PL} A, \\
(\textit{td-pcd}3) & [\textit{Triggered}^*_{\mathcal{I}}(\{C\}, \Gamma)]^f \vdash_{PL} A, \\
(\textit{td-pcd}3^+) & [\textit{Triggered}^*_{\mathcal{I}}(\{C\}, \Gamma)]^f \cup \{C\} \vdash_{PL} A, \\
(\textit{td-pcd}4) & \forall V \in \mathscr{L}_{PL}\bot\neg C : [\textit{Triggered}^*_{\mathcal{I}}(V, \Gamma)]^f \vdash_{PL} A, \\
(\textit{td-pcd}4^+) & \forall V \in \mathscr{L}_{PL}\bot\neg C : [\textit{Triggered}^*_{\mathcal{I}}(V, \Gamma)]^f \cup V \vdash_{PL} A.
\end{array}$$

So e.g. $(td - pcd1)$ defines A as obligatory if the truth of A is required to satisfy the triggered imperatives in any preferred subset. Of course, the crucial and as yet missing element is the decision procedure that determines the set $\mathscr{P}_{\mathcal{I}}(\{C\}, I)$ of preferred subsets. The next section discusses several proposals to define such subsets; a new proposal is presented in the section that follows it.

4 Identifying the preferred subsets

4.1 Brewka's preferred subtheories

The idea that normative conflicts can be overcome by use of a priority ordering of the norms involved dates back at least to Ross [35] and is also most prominent in von Wright's work (cf. [40] p. 68, 80).[14] However, it has turned out to be difficult to determine the exact mechanism by which such a resolution of conflicts can be achieved. This is true even when only unconditional imperatives are considered. Discussing various proposals for resolution of conflicts between unconditional imperatives, I have argued in [14] that an 'incremental' definition should be used for determining the relevant subsets. Based on earlier methods by Rescher [32], such a definition was first introduced by Brewka [3] for reasoning with prioritized

[14] But cf. already Watts [42] part II, ch. V, sec. III, principle 10: "Where two duties seem to stand in opposition to each other, and we cannot practise both, the less must give way to the greater, and the omission of the less is not sinful."

defaults. For any priority relation $<$, the idea is to consider all the 'full prioritizations' $\prec$ of $<$ (strict well orders that preserve $<$), and then work one's way from the top of the strict order downwards by adding the $\prec$-next-higher imperative to the thus constructed 'preferred subtheory' if its demand is consistent with the given facts and the demands of the imperatives that were added before. For the present setting, the definition can be given as follows:

Definition 1 (*Brewka's preferred subtheories*) Let $\mathcal{I} = \langle I, f, g, <\rangle$ be a prioritized conditional imperative structure, Δ be a subset of I, and $W \subseteq \mathscr{L}_{PL}$ be a set of *PL*-sentences. Then $\Gamma \in \mathscr{P}^{B}_{\mathcal{I}}(W, \Delta)$ iff (i) $W \nvdash_{PL} \bot$, and (ii) Γ is obtained from a full prioritization $\prec$ by defining

$$\Gamma_i = \begin{cases} \bigcup_{j \prec i} \Gamma_j \cup \{i\} & \text{if } W \cup \left[\bigcup_{j \prec i} \Gamma_j \cup \{i\}\right]^f \nvdash_{PL} \bot, \text{ and} \\ \bigcup_{j \prec i} \Gamma_j & \text{otherwise,} \end{cases}$$

for any $i \in \Delta$, and letting $\Gamma = \bigcup_{i \in \Delta} \Gamma_i$.

Clause (i) ensures that for an inconsistent set of assumed 'facts,' no set is preferred. Somewhat roundabout, owed to the possibility of infinite ascending subchains, clause (ii) then recursively defines a set $\Gamma \in \mathscr{P}^{\mathrm{B}}_{\mathcal{I}}(W, \Delta)$ for each full prioritization $\prec$: take the $\prec$-first i (the exclusion of infinite descending subchains guarantees that it exists) and if $W \cup \{i^f\} \nvdash_{PL} \bot$ then let $\Gamma_i = \{i\}$; otherwise Γ_i is left empty.[15] Similarly, any $\prec$-later i is tested for possible addition to the set $\bigcup_{j \prec i} \Gamma_j$ of elements that were added in the step for a $j \in \Delta$ that occurs $\prec$-prior to i. Γ is then the union of all these sets.

To see how this definition works, consider the set $I = \{!(p_1 \vee p_2), !\neg p_1, !\neg p_2\}$, with the ranking $!(p_1 \vee p_2) < !\neg p_2$ and $!\neg p_1 < !\neg p_2$. For an interpretation, let $!(p_1 \vee p_2))$ be your mother's request that you buy cucumbers or spinach for dinner, $!\neg p_1$ be your father's wish that no cucumbers are bought, and $!\neg p_2$ your sister's desire that you don't buy any spinach. We have two full prioritizations $!(p_1 \vee p_2) < !\neg p_1 < !\neg p_2$ and $!\neg p_1 < !(p_1 \vee p_2) < !\neg p_2$—let these be termed $\prec_1$ and $\prec_2$, respectively. The construction for $\prec_1$ adds the imperative $!(p_1 \vee p_2)$ in the first step and, since no conflict with the situation arises, $!\neg p_1$ in the second step. In the third and last step, nothing is added since $!\neg p_2$ conflicts with the demands of the already added imperatives. For $\prec_2$ the only difference is that the first two imperatives are added in inverse order. Thus $\mathscr{P}^{\mathrm{B}}_{\mathcal{I}}(W, I) = \{\{!(p_1 \vee p_2), !\neg p_1\}\}$. Using $(td - pcd1)$ we obtain $O(\neg p_1 \wedge p_2/\top)$, which means that you have to buy spinach and not cucumbers, thus fulfilling your parents' requests but not your sister's, which seems reasonable.

As I showed in [14], Brewka's method is extremely successful for dealing with unconditional imperatives. It is provably equivalent for such imperatives to methods proposed by Ryan [36] and Sakama and Inoue [37], and it avoids problems of other approaches by Alchourrón and Makinson [2], Prakken [30] and Prakken and Sartor [31]. Moreover, an equally intuitive maximization method proposed by Nebel [28], [29], that adds first a maximal set of the highest-ranking imperatives, then a maximal set of second-ranking imperatives, etc., but for its construction requires the ordering to be based on a complete preorder, can be shown to be embedded in Brewka's approach for such orderings. So my aim will be to retain Brewka's method for the unconditional case—in fact, all proposals that follow meet this criterion. However, when it is applied without change to conditional imperatives, the

[15] As usual, the union of an empty set of sets is taken to be the empty set.

algorithm may lead to incorrect results. E.g. consider a set I with two equally ranking imperatives $\{p_1 \Rightarrow !p_2, \neg p_1 \Rightarrow !\neg p_2\}$, meaning e.g. "if you go out, wear your boots" and "if you don't go out, don't wear your boots:" since the consequents contradict each other, an unmodified application of Brewka's method produces $\mathscr{P}^{\mathrm{B}}_{\mathcal{I}}(\{p_1\}, I) = \{\{p_1 \Rightarrow !p_2\}, \{\neg p_1 \Rightarrow !\neg p_2\}\}$, which fails to make true $O(p_2/p_1)$ by any truth definition of Sect. 3.4: the right set contains no imperatives that are triggered by p_1. So we cannot derive that you ought to wear your boots, given that you are going out. But intuitively there is no conflict, since the obligations arise in mutually exclusive circumstances only.

4.2 A naïve approach

A straightforward way to adopt Brewka's method to the case of conditional imperatives is to use not all imperatives for the construction, but only those that are triggered by the facts W, i.e. to use $\mathit{Triggered}_{\mathcal{I}}(W, \Delta)$ instead of Δ:

Definition 2 (*The naïve approach*) Let $\mathcal{I} = \langle I, f, g, <\rangle$ be a prioritized conditional imperative structure, Δ be a subset of I, and $W \subseteq \mathscr{L}_{PL}$ be a set of *PL*-sentences. Then $\Gamma \in \mathscr{P}^{n}_{\mathcal{I}}(W, \Delta)$ iff $\Gamma \in \mathscr{P}^{B}_{\mathcal{I}}(W, \mathit{Triggered}_{\mathcal{I}}(W, \Delta))$.

The change resolves our earlier problems with Brewka's method: consider again the set of imperatives $\{p_1 \Rightarrow !p_2, \neg p_1 \Rightarrow !\neg p_2\}$, where the imperatives were interpreted as ordering me to wear my boots when I go out, and not wear my boots when I don't. The new definition produces $\mathscr{P}^{\mathrm{n}}_{\mathcal{I}}(\{p_1\}, I) = \{\{p_1 \Rightarrow !p_2\}\}$, its only 'preferred' subset containing just the one imperative that is triggered given the facts $\{p_1\}$. By all truth definitions of Sect. 3.4, $O(p_2/p_1)$ is now true, so given that you go out, you ought to wear your boots, which is as it should be.

The naïve approach is similar to Horty's proposal in [20] in that conflicts are only removed between imperatives that are triggered (though the exact mechanism differs from Horty's). When I nevertheless call it 'naïve,' this is because there are conceivable counterexamples to this method. Consider the set of prioritized imperatives $!p_1 < p_1 \Rightarrow !p_2 < !\neg p_2$, and for an interpretation suppose that your job requires you to go outside p_1, that your mother, who is concerned for your health, told you to wear a scarf p_2 if you go outside, and that your friends don't want you to wear a scarf, whether you go outside or not. In the default situation $\top$ only the first imperative and the third are triggered, i.e. $\mathit{Triggered}_{\mathcal{I}}(\{\top\}, I) = \{!p_1, !\neg p_2\}$. Since their demands are consistent with each other, we obtain $\mathscr{P}^{\mathrm{n}}_{\mathcal{I}}(\{\top\}, I) = \{\{!p_1, !\neg p_2\}\}$, for which all truth definitions of Sect. 3.4 make $O(p_1 \wedge \neg p_2/\top)$ true. So you ought to go out and not wear a scarf, thus satisfying the first and the third imperative, but violating the second-ranking imperative. But arguably, if an imperative is to be violated, it should not be the second-ranking $p_1 \Rightarrow !p_2$, but the lowest ranking $!\neg p_2$ instead.

4.3 The stepwise approach

To avoid the difficulties of the 'naïve' approach, it seems we must not just take into account the imperatives that are triggered, but also those that become triggered when higher ranking imperatives are satisfied. To this effect, the following modification may seem reasonable:

Definition 3 (*The stepwise approach*) Let $\mathcal{I} = \langle I, f, g, <\rangle$ be a prioritized conditional imperative structure, Δ be a subset of I, and $W \subseteq \mathscr{L}_{PL}$ be a set of *PL*-sentences. Then

$\Gamma \in \mathscr{P}^s(W, \Delta)$ iff (i) $W \nvdash_{PL} \bot$, and (ii) Γ is obtained from a full prioritization $\prec$ by defining

$$\Gamma_i = \begin{cases} \bigcup_{j \prec i} \Gamma_j \cup \{i\} & \text{if } i \in \mathit{Satisfiable}_{\mathcal{I}}(W \cup \left[\bigcup_{j \prec i} \Gamma_j\right]^f, \Delta), \text{ and} \\ \bigcup_{j \prec i} \Gamma_j & \text{otherwise,} \end{cases}$$

for any $i \in \Delta$, and letting $\Gamma = \bigcup_{i \in \Delta} \Gamma_i$.

So at each step one considers what happens if the imperatives that were included so far are satisfied, and adds the current imperative only if it is satisfiable given W and the satisfaction of these previous imperatives. Note that satisfiability of an imperative, like its satisfaction and violation, presupposes that the imperative is triggered. The new definition not only includes, at each step, those imperatives that are triggered and can be satisfied given the facts and the supposed satisfaction of the previously added imperatives: it also includes those that *become* triggered when a previously added imperative is satisfied.

This modification avoids the previous difficulty: consider again the set of prioritized imperatives $!p_1 < p_1 \Rightarrow !p_2 < !\neg p_2$. There is just one full prioritization, which for $W = \{\top\}$ yields in the first step the set $\{!p_1\}$, and in the second step $\{!p_1, p_1 \Rightarrow !p_2\}$, since $p_1 \Rightarrow !p_2$ is triggered when the previously added imperative $!p_1$ is assumed to be satisfied. In the third step, nothing is added: though the imperative $!\neg p_2$ is triggered, it cannot be satisfied together with the previously added imperatives. So we obtain $\mathscr{P}^s_{\mathcal{I}}(\{\top\}, I) = \{\{!p_1, p_1 \Rightarrow !p_2\}\}$, and hence $O(p_1/\top)$, but not $O(p_1 \wedge \neg p_2/\top)$, is defined true by all truth definitions of Sect. 3.4. Operators that accept 'deontic detachment' (*td-pcd*$3^{(+)}, 4^{(+)}$) make true $O(p_1 \wedge p_2/\top)$, so you must go out and wear a scarf, which is as it should be.

However, a small change in the ordering shows that this definition does not suffice: let the imperatives now be ranked $p_1 \Rightarrow !p_2 < !p_1 < !\neg p_2$. (For the interpretation, assume that the conditional imperative to wear a scarf when leaving the house was uttered by some high-ranking authority, e.g. a doctor.) Then again $\mathscr{P}^s_{\mathcal{I}}(\{\top\}, I) = \{\{!p_1, !\neg p_2\}\}$: in the first step, nothing is added since $p_1 \Rightarrow !p_2$ is neither triggered by the facts nor by the assumed satisfaction of previously added imperatives (there are none). In the next two steps, $!p_1$ and $!\neg p_2$ are added, as each is consistent with the facts and the satisfaction of the previously added imperatives. So again all truth definitions of Sect. 3.4 make true $O(p_1 \wedge \neg p_2/\top)$, i.e. you ought to go out and not wear a scarf, satisfying the second and third ranking imperatives at the expense of the highest ranking one. But surely, if one must violate an imperative, it should be one of the lower-ranking ones instead.

4.4 The reconsidering approach

The merits of the stepwise approach were that it did not only consider the imperatives that are triggered, but also those that *become* triggered when already added imperatives are satisfied. Such considerations applied to those imperatives that follow in the procedure. Yet the satisfaction of already added imperatives might also trigger higher-ranking imperatives, which by this method are not considered again. So it seems necessary, at each step, to reconsider also the higher-ranking imperatives. An algorithm that does that was first introduced for default theory by Marek and Truszczyński [27] p. 72, and later employed by Brewka in [4]; it can be reformulated for the present setting as follows:

Definition 4 (*The reconsidering approach*) Let $\mathcal{I} = \langle I, f, g, < \rangle$ be a prioritized conditional imperative structure, Δ be a subset of I, and $W \subseteq \mathscr{L}_{PL}$ be a set of *PL*-sentences. Then

$\Gamma \in \mathscr{P}^{r}_{\mathcal{I}}(W, \Delta)$ iff (i) $W \nvdash_{PL} \bot$, and (ii) Γ is obtained from a full prioritization $\prec$ by defining

$$\Gamma_i = \bigcup\nolimits_{j \prec i} \Gamma_j \cup min_{\prec}\left[\mathit{Satisfiable}_{\mathcal{I}}\left(W \cup \left[\bigcup\nolimits_{j \prec i} \Gamma_j\right]^f, \Delta\right) \setminus \bigcup\nolimits_{j \prec i} \Gamma_j\right]$$

for $i \in \Delta$, and letting $\Gamma = \bigcup_{i \in \Delta} \Gamma_i$.[16]

The definition reconsiders at each step the whole ordering, and adds the $\prec$-first[17] imperative (due to the definition of $\prec$ there is just one) that has not been added previously and is satisfiable given both the facts W and the consequents of the previously added imperatives. To see how the definition works, consider again the example which the stepwise approach failed, i.e. the set of prioritized imperatives $p_1 \Rightarrow !p_2 < !p_1 < !\neg p_2$. We are interested in the preferred sets for the default circumstances $\top$, i.e. the sets in $\mathscr{P}^{r}_{\mathcal{I}}(\{\top\}, I)$. I is already fully prioritized, so there is just one such set. Applying the algorithm, we find the minimal (highest ranking) element in $\mathit{Satisfiable}_{\mathcal{I}}(\{\top\}, I)$ is $!p_1$, so this element is added in the first step. In the second step, we look for the minimal element in $\mathit{Satisfiable}_{\mathcal{I}}(\{\top\} \cup \{!p_1\}^f, I)$, other than the previously added $!p_1$. It is $p_1 \Rightarrow !p_2$, since the assumed satisfaction of all previously added imperatives triggers it, and its consequent can be true together with $\{\top\} \cup \{p_1\}$. So $p_1 \Rightarrow !p_2$ is added in this step. In the remaining third step, nothing is added: $!\neg p_2$ is not in $\mathit{Satisfiable}_{\mathcal{I}}(\{\top\} \cup \{!p_1, p_1 \Rightarrow !p_2\}^f, I)$, and all other elements in this set have been previously added. So $\mathscr{P}^{r}_{\mathcal{I}}(\{\top\}, I) = \{\{!p_1, p_1 \Rightarrow !p_2\}\}$. Now all truth definitions of Sect. 3.4 make true $O(p_1/\top)$, but not $O(p_1 \wedge \neg !p_2/\top)$, and operators that accept 'deontic detachment' make true $O(p_1 \wedge p_2/\top)$. So, in the given interpretation, you must go out (as your job requires) and wear a scarf (as the doctor ordered you to do in case you go out), which is as it should be.

However, again problems remain. Let the imperatives now be prioritized in the order $p_1 \Rightarrow !p_2 < !\neg p_2 < !p_1$. Let $p_1 \Rightarrow !p_2$ stand for the doctor's order to wear a scarf when going outside, let $!\neg p_2$ stand for your friends' expectation that you don't wear a scarf, and let $!p_1$ represent your sister's wish that you leave the house. Construct the set in $\mathscr{P}^{r}_{\mathcal{I}}(\{\top\}, I)$—since I remains fully prioritized, there is again just one such set. The minimal element in $\mathit{Satisfiable}_{\mathcal{I}}(\{\top\}, I)$ is $!\neg p_2$, and so is added in the first step. The minimal element in $\mathit{Satisfiable}_{\mathcal{I}}(\{\top\} \cup \{!\neg p_2\}^f, I)$, other than $!\neg p_2$, is $!p_1$ which therefore gets added in the second step. Nothing is added in the remaining step: $!\neg p_2$ and $!p_1$ have already been added, and $p_1 \Rightarrow !p_2$ is not in $\mathit{Satisfiable}_{\mathcal{I}}(\{\top\} \cup \{!\neg p_2, !p_1\}^f, I)$: though it is triggered by the assumed satisfaction of $!p_1$, its consequent is contradicted by the assumed satisfaction of $!\neg p_2$. So $\mathscr{P}^{r}_{\mathcal{I}}(\{\top\}, I) = \{\{!p_1, !\neg p_2\}\}$. Hence all truth definitions of Sect. 3.4 again make true $O(p_1 \wedge \neg p_2/\top)$, so you ought to go out without a scarf, again satisfying the second and third ranking imperatives at the expense of the first, which seems the wrong solution.

4.5 A fixpoint approach

To eliminate cases in which the 'reconsidering approach' still adds imperatives that can only be satisfied at the expense of violating a higher-ranking one, a 'fixpoint' approach was first proposed for default reasoning by Brewka and Eiter [5]. It tests each set that may be considered as preferred to see if it really includes all the elements that should be included:

[16] Note that in 'Γ_i,' i is used just as an index—it does not mean that i is considered for addition at this step, and may be added at an earlier or later step (or not at all).

[17] For any ordering $<$ on some set Γ, $min_{<}\Gamma = \{i \in \Gamma | \forall i' \in \Gamma : \text{if } i' \neq i, \text{then } i' \nless i\}$, and $max_{<}\Gamma = \{i \in \Gamma | \forall i' \in \Gamma : \text{if } i' \neq i, \text{then } i \nless i'\}$, as usual.

imperatives that are triggered given the facts and the assumed satisfaction of all imperatives in the set, and would be added by Brewka's [3] original method that adds the higher ranking imperatives first. The procedure translates as follows:

Definition 5 (*The fixpoint approach*) Let $\mathcal{I} = \langle I, f, g, < \rangle$ be a prioritized conditional imperative structure, Δ be a subset of I, and $W \subseteq \mathscr{L}_{PL}$ be a set of *PL*-sentences. Then

$$\Gamma \in \mathscr{P}^{f}_{\mathcal{I}}(W, \Delta) \quad \text{iff} \quad \Gamma \in \mathscr{P}^{B}_{\mathcal{I}}(W, \mathit{Triggered}_{\mathcal{I}}(W \cup \Gamma^{f}, \Delta)).$$

To see how this definition works, consider the above set of prioritized imperatives $p_1 \Rightarrow !p_2 < !\neg p_2 < !p_1$. It is immediate that the set $\{!p_1, !\neg p_2\}$ cannot be in $\mathscr{P}^{\mathrm{f}}_{\mathcal{I}}(\{\top\}, I)$: if we assume all imperatives in this set to be satisfied, then all imperatives are triggered, i.e. $\mathit{Triggered}_{\mathcal{I}}(\{\top\} \cup \{!p_1, !\neg p_2\}^{f}, I) = I$. By Brewka's original method, $\mathscr{P}^{\mathrm{B}}_{\mathcal{I}}(W, I) = \{\{p_1 \Rightarrow !p_2, !p_1\}\}$: $<$ is already fully prioritized, and for this full prioritization the method adds $p_1 \Rightarrow !p_2$ in the first step, $!\neg p_2$ cannot be added in the second step since its consequent contradicts the consequent of the previously added $p_1 \Rightarrow !p_2$, and in the third step $!p_1$ is added. So since the considered set $\{!p_1, !\neg p_2\}$ is not in $\mathscr{P}^{\mathrm{B}}_{\mathcal{I}}(W, I)$, it is not a 'fixpoint.' Rather, as may be checked, the only 'fixpoint' in $\mathscr{P}^{\mathrm{f}}_{\mathcal{I}}(\{\top\}, I)$ is $\{p_1 \Rightarrow !p_2, !p_1\}$. Then all truth definitions of Sect. 3.4 make true $O(p_1/\top)$, but no longer $O(p_1 \wedge \neg p_2/\top)$. Moreover, truth definitions that allow 'deontic detachment' make true $O(p_1 \wedge p_2/\top)$. In the given interpretation this means that you must leave the house at your sisters request and wear a scarf, as the doctor ordered you to do in case you go out.

Though the construction now no longer makes true $O(p_1 \wedge \neg p_2/\top)$, its solution for the example, that determines the set $\{p_1 \Rightarrow !p_2, !p_1\}$ as the fixpoint of the set of prioritized imperatives $p_1 \Rightarrow !p_2 < !\neg p_2 < !p_1$, seems questionable. Though this now includes the doctor's order, you now have no obligation anymore to satisfy the imperative that is second ranking, i.e. your friends' request that you don't wear a scarf; truth definitions that accept 'deontic detachment' even oblige you to violate it by wearing a scarf. Now consider the situation without the third ranking imperative $!p_1$: it can easily be verified that for a set $I = \{p_1 \Rightarrow !p_2, !\neg p_2\}$ the only fixpoint in $\mathscr{P}^{\mathrm{f}}_{\mathcal{I}}(\{\top\}, I)$ is $\{!\neg p_2\}$. So for the reduced set, $(td - pcd1)$ makes true $O(\neg p_2/\top)$, i.e. you ought to obey your friends' wish. That the satisfaction of this higher ranking imperative $!\neg p_2$ should no longer be obligatory when a lower ranking imperative $!p_1$ is added, seems hard to explain. If the ranking is taken seriously, I think one should still satisfy the higher ranking imperatives, regardless of what lower ranking imperatives are added.

For another problem consider the set of prioritized imperatives $p_1 \Rightarrow !p_2 < !(p_1 \wedge \neg p_2) < !p_3$. For an interpretation, let the first imperative be again the doctor's order to wear a scarf in case you go out, the second one be your friends' request to go out and not wear a scarf, and the third be the wish of your aunt that you write her a letter. It is easily proved that the set has no fixpoint, and so there is also none that contains $!p_3$, hence all truth definitions make $O(p_3/\top)$ false, so you do not even have to write to your aunt. But even if the presence of a higher ranking conditional imperative and a lower ranking imperative to violate it poses a problem (why should it? after all, the lower ranking imperative is outranked), it is hard to see why the subject should be left off the hook for all other, completely unrelated obligations.[18]

[18] The lack of fixpoints is a well-known problem of such definitions (cf. e.g. Caminada and Sakama [7]). Another approach to conditional imperatives by Makinson in [24] has trouble resolving the same example: for the default circumstances $\top$ it produces the set $\{!(p_1 \wedge \neg p_2), !p_3\}$. $p_1 \Rightarrow !p_2$ is not considered, since its only 'label' (roughly: a conjunction of the circumstances, the imperatives' antecedents that would trigger* it, and its consequent) is inconsistent (it is $\top \wedge (p_1 \wedge \neg p_2) \wedge p_2$). But why should the agent not be free to obey $p_1 \Rightarrow !p_2$, and not violate it by satisfying $!(p_1 \wedge \neg p_2)$?

4.6 Discussion

For a discussion of our results so far, let us return to the ‘drinking and driving’ example from the introduction. Let the three imperatives:

(1) Your mother says: if you drink anything, then don’t drive.
(2) Your best friend says: if you go to the party, then you do the driving.
(3) Some acquaintance says: if you go to the party, then have a drink with me.

be represented by the set of prioritized imperatives $p_1 \Rightarrow \neg p_2 < p_3 \Rightarrow p_2 < p_3 \Rightarrow p_1$. Let the set of facts be $\{p_3\}$, i.e. you go to the party. Brewka’s original method is not tailored to be directly employed on conditional imperatives, as it ignores the antecedents altogether. The next three approaches, the naïve, the stepwise and the reconsidering ones, produce $\mathscr{P}^{\mathrm{n}}_{\mathcal{I}}(\{p_3\}, I) = \mathscr{P}^{\mathrm{s}}_{\mathcal{I}}(\{p_3\}, I) = \mathscr{P}^{\mathrm{r}}_{\mathcal{I}}(\{p_3\}, I) = \{\{p_3 \Rightarrow !p_2, p_3 \Rightarrow !p_1\}\}$, which by all truth definitions of Sect. 3.4 makes true $O(p_1 \wedge p_2/p_3)$, so you ought to drink and drive. The fixpoint approach produces $\mathscr{P}^{\mathrm{f}}_{\mathcal{I}}(\{p_3\}, I) = \{\{p_1 \Rightarrow !\neg p_2, p_3 \Rightarrow !p_1\}\}$, so all truth definitions make true $O(p_1/p_3)$, which means you ought to drink. Truth definitions with ‘deontic detachment’ additionally make true $O(p_1 \wedge \neg p_2/p_3)$, so you ought to drink and not drive. But being obliged to drink runs counter to our intuitions for the ‘drinking and driving’ example. So we have to look for a different solution.

Before we do that, I will, however, question again our intuition in this matter. Horty [21] has recently used a structurally identical example to argue for just the opposite, that the solution by the fixpoint approach is correct. His example is that of three commands, uttered by a colonel, a major and a captain to a soldier, Corporal O’Reilly. The Colonel, who does not like it too warm in the cabin, orders O’Reilly to open the window whenever the heat is turned on. The Major, who is a conservationist, wants O’Reilly to keep the window closed during the winter. And the Captain, who does not like it to be cold, orders O’Reilly to turn the heat on during the winter. O’Reilly is trying to figure out what to do. The intended representation is again the prioritized conditional imperative structure employed above for the ‘drinking and driving’ example, where p_1 now means that the heat is turned on, p_2 means that the window is closed, and p_3 means that it is winter. We saw that the fixpoint approach yields the preferred subset $\{p_1 \Rightarrow !\neg p_2, p_3 \Rightarrow !p_1\}$, making true $O(p_1/p_3)$ for all truth definitions, and $O(p_1 \wedge \neg p_2/p_3)$ for truth definitions that accept ‘deontic detachment,’ so O’Reilly must turn on the heat and then open the window, and thus violate the Major’s order. Horty argues as follows in support of this choice:

> “O’Reilly’s job is to obey the orders he has been given exactly as they have been issued. If he fails to obey an order issued by an officer without an acceptable excuse, he will be court-martialed. And, let us suppose, there is only one acceptable excuse for failing to obey such an order: that obeying the order would, in the situation, involve disobeying an order issued by an officer of equal or higher rank. (...) So given the set of commands that O’Reilly has been issued, can he, in fact, avoid court-martial? Yes he can, by (...) obeying the orders issued by the Captain and the Colonel (...). O’Reilly fails to obey the Major’s order, but he has an excuse: obeying the Major’s order would involve disobeying an order issued by the Colonel.”

Horty’s principle seems quite acceptable: for each order issued to the agent, the agent may ask herself if obeying the order would involve disobeying an order of a higher ranking officer (then she is excused), and otherwise follow it. The result is a set of imperatives where each imperative is either obeyed, or disobeyed but the disobedience excused. When I nevertheless think the argument is not correct, it is because I think it confuses the *status quo* and the *status* *quo posterior*. Obeying the Major's order does not, in the initial situation, involve disobeying the Colonel's order. Only once O'Reilly follows the Captain's order and turns on the heat, it is true that he must obey the Colonel, open the window and thus violate the Major's order. But this does not mean that he should follow the Captain's order in the first place—as by doing so he brings about a situation in which he is forced, by a higher ranking order, to violate a command from another higher ranking officer. Quite to the contrary, I think that being forced to violate a higher ranking order when obeying a lower ranking one is a case where following the lower one 'involves' such a violation, and so the only order the agent is excused from obeying is the lowest ranking command.

Consider finally this variant: suppose that if I am attacked by a man, I must fight him (to defend my life, my family etc.). Furthermore, suppose I have pacifist ideals which include that I must not fight the man. Now you tell me to provoke him, which in the given situation means that he will attack me. Let self-defense rank higher than my ideals, which in turn rank higher than your request. Should I do as you request? By the reasoning advocated by Horty, there is nothing wrong with it: I satisfy your request, defend myself as I must, and though I violate my ideals, I can point out to myself that the requirement to fight back took priority. But I think if I really do follow your advice, I would feel bad. I think this would not just be some irrational regret for having to violate, as I must, my ideals, but true guilt for having been tempted into doing something I should not have done, namely provoking the man: it caused the situation that made me violate my ideals. So I think our intuitions in the 'drinking and driving' example and the other cases have been correct.

5 New strategies and a new proposal

In the face of the difficulties encountered so far, it seems necessary to address the issue of finding an appropriate mechanism for a resolution of conflicts between prioritized conditional imperatives in a more systematic manner. So far intuition has guided us mainly as to what imperatives should be included in a subset of 'preferred imperatives.' I think the following postulate sums up the intuitions that have so far influenced the proposal and rejection of solutions:

Postulate *Any imperative should be considered relevant (included) as long as it is not violated or, in the given situation, conflicts with other imperatives that are also considered relevant (included) and do not rank lower.*

The postulate makes clear that the need to satisfy a lower ranking imperative cannot serve as an excuse to violate an imperative of higher priority. However, this postulate delegates the answer of how the set of 'preferred imperatives' should be constructed to the answer of another question: when do conditional imperatives conflict in a given situation? There appear to be several possible answers to this question, which lead to different solutions.

5.1 Subsets with consistent extensions

For the definition of conflicts between conditional imperatives one might recur to a definition of conflicts between desires as proposed by Broersen et al. [6]. Their idea, translated to our setting, is that for any set of facts W and set of conditional imperatives Δ there is a smallest set $E(W, \Delta) \subseteq \mathscr{L}_{PL}$ such that (i) $W \subseteq E(W, \Delta)$ and (ii) for any $i \in \Delta$, if $E(W, \Delta) \vdash_{PL} g(i)$ then $f(i) \in E(W, \Delta)$. The set of imperatives Δ is then defined as conflicting given the facts W if the thus constructed extension $E(W, \Delta)$ of the facts is inconsistent, i.e. $E(W, \Delta) \vdash_{PL} \bot$

(cf. [6] def. 5.5). This 'test' for conflicts may then be employed within a variant of Brewka's original method.

Definition 6 (*Preferred Subsets with Consistent Extensions*) Let $\mathcal{I} = \langle I, f, g, < \rangle$ be a prioritized conditional imperative structure, Δ be a subset of I, and $W \subseteq \mathscr{L}_{PL}$ be a set of *PL*-sentences, and let $E(W, \Delta)$ be defined as described above. Then $\Gamma \in \mathscr{P}^e_{\mathcal{I}}(W, \Delta)$ iff (i) $W \nvdash_{PL} \bot$, and (ii) Γ is obtained from a full prioritization $\prec$ by defining

$$\Gamma_i = \begin{cases} \bigcup_{j \prec i} \Gamma_j \cup \{i\} & \text{if } E(W, \bigcup_{j \prec i} \Gamma_j \cup \{i\}) \nvdash_{PL} \bot, \text{ and} \\ \bigcup_{j \prec i} \Gamma_j & \text{otherwise,} \end{cases}$$

for any $i \in \Delta$, and letting $\Gamma = \bigcup_{i \in \Delta} \Gamma_i$.

To see how this definition works, consider the 'drinking and driving' example, where the set of prioritized imperatives is $p_1 \Rightarrow !\neg p_2 < p_3 \Rightarrow !p_2 < p_3 \Rightarrow !p_1$ and the situation $W = \{p_3\}$. There is only one full prioritization which is identical with $<$. In the first step, $p_1 \Rightarrow !\neg p_2$ is added, since $E(\{p_3\}, \{p_1 \Rightarrow !\neg p_2\}) = \{p_3\}$ which is consistent. In the second step, $p_3 \Rightarrow !p_2$ is added, as $E(\{p_3\}, \{p_1 \Rightarrow !\neg p_2, p_3 \Rightarrow !p_2\}) = \{p_3, p_2\}$, which is again consistent. In the third step, $p_3 \Rightarrow !p_1$ is rejected, since $E(\{p_3\}, \{p_1 \Rightarrow !\neg p_2, p_3 \Rightarrow !p_2, p_3 \Rightarrow !p_1\}) = \{p_3, p_2, p_1, \neg p_2\}$, which is inconsistent. So we have $\mathscr{P}^e_{\mathcal{I}}(\{p_3\}, I) = \{\{p_1 \Rightarrow !\neg p_2, p_3 \Rightarrow !p_2\}\}$, making true $O(p_2/p_3)$ for all truth definitions of Sect. 3.4, so given that I go to the party I must do the driving, which is as it should be.

However, there is a problem for the test using consistent extensions, as for some truth definitions it delivers sets of imperatives that are clearly conflicting, in the sense that they make $O(\bot/C)$ true for a consistent fact C: Consider the set of facts $W = \{p_1 \vee p_2\}$ and the set of imperatives $I = \{p_1 \Rightarrow !p_3, p_2 \Rightarrow !p_3, !\neg p_3\}$. We have $E(W, I) = \{p_1 \vee p_2, \neg p_3\}$, which is consistent, and so all imperatives in I are added to the preferred subset, regardless of their ordering. But for any truth definition that allows for 'reasoning by cases' (*td-pcd*$2^{(+)}, 4^{(+)}$) we then have both $O(p_3/p_1 \vee p_2)$ and $O(\neg p_3/p_1 \vee p_2)$, and so also $O(\bot/p_1 \vee p_2)$. This is simply because the construction of extensions does not take care of reasoning by cases, i.e. it does not add p_3 to the extension in case we both have $p_1 \vee p_2$ in W and $\{p_1 \Rightarrow !p_3, p_2 \Rightarrow !p_3\} \subseteq I$. Perhaps the definition of consistent extensions can be amended, but to avoid delivering preferred sets of imperatives that make true $O(\bot/C)$ for some truth definition and some consistent fact C suggests a different solution that is explored in the next section.

5.2 Deontically tailored preferred subsets

In the unconditional case, the reason to move from definition (*td-m*1) to (*td-m*2) was that when there are conflicts between imperatives, the former makes true the monadic deontic formula $O\bot$, i.e. the agent ought to do the logically impossible. This result was avoided by considering only maximal sets of imperatives with demands that are collectively consistent, i.e. sets that do not make $O\bot$ true. When faced with the question what dyadic deontic formula should not be true when conflicts are resolved for arbitrary situations C, the formula $O(\neg C/C)$ appears to be the dyadic equivalent: a mechanism for conflict resolution should not result in telling the agent to change the supposed, unalterable facts.[19] So to define the set $\mathscr{P}_{\mathcal{I}}(\{C\}, I)$ required by the truth definitions (*td-pcd*$1^{(+)} - 4^{+}$), we can modify Brewka's

[19] This test is identical to the one used by Makinson and van der Torre [26] p. 158/159 to determine 'consistency of output' (cf. also for arguments why $O(\neg C/C)$ should be used, i.e. for their setting, the 'output' should be consistent with the 'input,' rather than the formula $O(\bot/C)$ and thus consistency of output *simpliciter*).

original method in such a way that it tests, at each step, for each of the constructed subsets, if the corresponding truth-definition (*td-cd*$1^{(+)} - 4^{+}$) does not make $O(\neg C/C)$ true for this set.[20] Formally:

Definition 7 (*Deontically Tailored Preferred Subsets*) Let $\mathcal{I} = \langle I, f, g, < \rangle$ be a prioritized conditional imperative structure, and $C \in \mathscr{L}_{PL}$ describe the given situation. Let (*td-pcd*$*$) be any of the truth definitions (*td-pcd*$1^{(+)} - 4^{(+)}$). Then Γ is in the set $\mathscr{P}^{*}_{\mathcal{I}}(\{C\}, I)$ employed by this truth definition iff (i) $\{C\} \nvdash_{PL} \bot$, and (ii) Γ is obtained from a full prioritization $\prec$ by defining

$$\Gamma_i = \begin{cases} \bigcup_{j \prec i} \Gamma_j \cup \{i\} & \text{if } \langle \bigcup_{j \prec i} \Gamma_j \cup \{i\}, f, g \rangle \nvDash O(\neg C/C) by (\textit{td-cd}*), \\ \bigcup_{j \prec i} \Gamma_j & \text{otherwise,} \end{cases}$$

for any $i \in I$, and letting $\Gamma = \bigcup_{i \in I} \Gamma_i$.

By this construction, each of the preferred subsets contains a maximal number of the imperatives such that they do not make true $O(\neg C/C)$ for the situation C and the truth definition that is employed, and so the resulting truth definition likewise avoids this truth. Such a construction of the preferred subsets might be considered 'tailored' to the truth definition in question, and any remaining deficiencies might be seen as stemming from the employed truth definition. But this being so, the method reveals a strong bias towards truth definitions that accept 'deontic detachment', and in particular truth definitions (*td-pcd*$2^{+}, 3^{(+)}, 4^{(+)}$):

Consider the set of imperatives $I = \{!p_1, p_1 \Rightarrow !p_2, !\neg p_2\}$ with the ranking $!p_1 < p_1 \Rightarrow !p_2 < !\neg p_2$, that was used to refute the 'naïve approach.' As can be easily checked, $\mathscr{P}^{*}_{\mathcal{I}}(\{\top\}, I) = \{I\}$ for all truth definitions (*td-pcd*$1, 1^{+}, 2$). So by all these truth definitions, $O(p_1 \wedge \neg p_2/\top)$ is true. So they commit us to violating the second-ranking imperative, whereas intuitively, the third-ranking imperative should be violated instead. By contrast, all truth definitions (*td-pcd*$3^{(+)}, 4^{(+)}$), that employ reusable output, and of course likewise (*td-pcd*2^{+}) that is equivalent to (*td-pcd*4^{+}), handle all given examples exactly as it was suggested they should. In particular, consider the ordered imperatives $p_1 \Rightarrow !p_2 < !\neg p_2 < !p_1$, that were used to refute both the 'reconsidering' and the 'fixpoint' approaches: for $* = 2^{+}, 3^{(+)}, 4^{(+)}$, $\mathscr{P}^{*}_{\mathcal{I}}(\{\top\}, I)$ is $\{\{p_1 \Rightarrow !p_2, !\neg p_2\}\}$, making $O(\neg p_2/\top)$ true by all these truth definitions, which thus commit us to satisfying the second ranking imperative, and not to violating it in favor of satisfying the third ranking imperative as these approaches did. The 'drinking and driving' example is also handled correctly: the set of prioritized imperatives $p_1 \Rightarrow !\neg p_2 < p_3 \Rightarrow !p_2 < p_3 \Rightarrow !p_1$ produces, for the situation p_3, the set $\mathscr{P}^{*}_{\mathcal{I}}(\{p_3\}, I) = \{\{p_1 \Rightarrow !\neg p_2, p_3 \Rightarrow !p_2\}\}$. So the third ranking imperative, that commits the agent to drinking and thus, by observation of the highest ranking imperative, prevents the agent from driving, is disregarded. Instead, the truth definitions make true $O(p_2/p_3)$, so the agent must do the driving if she goes to the party, as her best friend asked her to.

Is this the solution, then? Some uneasiness remains as to the quick way with which definitions (*td-pcd*$1, 1^{+}, 2$) are discharged as insufficient. Why should it not be possible to maintain, as these definitions do, that conditional imperatives only produce an obligation if they are factually triggered, while at the same time maintaining that the above examples should not be resolved the way they are? The purpose of a truth definition for the deontic O-operator is to find a formal notion of 'ought' that reflects ordinary reasoning, and our intuitions on that matter may differ from our ideas as to what may constitute a good choice from a possibly conflicting set of prioritized conditional imperatives. I will now make a new proposal how to construct the 'preferable subsets', that keeps the positive results without

[20] The preferred subsets are thus a choice from the 'maxfamilies' defined in [26].

committing us to prefer any of the truth definitions of Sect. 3.4 by virtue of their handling of prioritized imperatives alone.

5.3 Preferred maximally obeyable subsets

What made Brewka's approach so successful is that it maximizes the number of higher ranking imperatives in the preferred subsets of a given set of unconditional imperatives: for each 'rank,' a maximal number of imperatives are added that can be without making the set's demands inconsistent in the given situation. As was shown, Brewka's approach cannot be directly applied to conditional imperatives, since it makes no sense to test the demands of imperatives for inconsistencies if these imperatives may not be triggered in the same circumstances. Just considering triggered imperatives is also not enough, as was demonstrated for the 'naïve approach.' But if the maximization method is to include imperatives that are not (yet) triggered, then we must look for something else than inconsistency of demands to take the role of a threshold criterion for the maximization process.

To do so we should ask ourselves why, for the unconditional case, the aim was to find a maximal set of imperatives with demands that are collectively consistent with the situation. I think that by doing so we intend to give the agent directives that can be safely followed. While in the unconditional case this means that the agent can satisfy all the chosen imperatives, the situation is different for conditional imperatives: here an agent can also obey imperatives without necessarily satisfying their demands. If you tell me to visit you in case I go to Luxembourg next month, I can safely arrange to spend all of next month at home and still do nothing wrong. If we think not so much of imperatives, but of legal regulations, then I can obviously be a law-abiding citizen by simply failing to trigger any legal norm (even though this might imply living alone on an island): whether I do that or boldly trigger some of the regulations' antecedents and then satisfy those I have triggered seems not a question of logic, but of individual choice. So I think the threshold criterion to be used should be that of obeyability: we should maximize the set of imperatives the agent can obey, and only disregard an imperative if its addition to the set means that at least one of the added imperatives must (now) be violated, given the facts.[21]

For a given set of conditional imperatives Δ and a set of factual truths W, the subsets of imperatives that can be obeyed are described by $\mathit{Obeyable}_{\mathcal{I}}(W, \Delta)$, i.e. they are those subsets $\Gamma \subseteq \Delta$ such that $W \cup \Gamma^m \nvdash_{PL} \bot$. To maximize not by collective consistency of demands, but by collective obeyability, Brewka's original approach can therefore be changed as follows:

Definition 8 (*Preferred Maximally Obeyable Subsets*) Let $\mathcal{I} = \langle I, f, g, < \rangle$ be a prioritized conditional imperative structure, Δ be a subset of I, and $W \subseteq \mathscr{L}_{PL}$ be a set of *PL*-sentences. Then $\Gamma \in \mathscr{P}^o_{\mathcal{I}}(W, \Delta)$ iff (i) $W \nvdash_{PL} \bot$, and (ii) Γ is obtained from a full prioritization $\prec$ by defining

$$\Gamma_i = \begin{cases} \bigcup_{j \prec i} \Gamma_j \cup \{i\} & \text{if } \bigcup_{j \prec i} \Gamma_j \cup \{i\} \in \mathit{Obeyable}_{\mathcal{I}}(W, \Delta), \text{ and} \\ \bigcup_{j \prec i} \Gamma_j & \text{otherwise,} \end{cases}$$

for any $i \in \Delta$, and letting $\Gamma = \bigcup_{i \in \Delta} \Gamma_i$.

The change from Brewka's original definition is only minute: we test not the demands of the imperatives for consistency, but their materializations. Note that this is a conservative extension of Brewka's method, since for any unconditional imperative i we have $\vdash_{PL} f(i) \leftrightarrow$

[21] While Hansson, in [16] p. 200, also advocates a move from 'consistency' to 'obeyability,' what is meant there is rather the step from (*td-m*2) to (*td-d*1).

$m(i)$. As can easily be seen, the new construction solves all of the previously considered difficulties, regardless of the chosen truth definition for the deontic O-operator:

- To refute a direct application of Brewka's original method, we used the set $I = \{p_1 \Rightarrow p_2, \neg p_1 \Rightarrow \neg p_2\}$ with no ranking imposed. $m(I)$ is consistent and so $\mathscr{P}^{o}_{\mathcal{I}}(\{p_1\}, I) = \{I\}$, making $O(p_2/p_1)$ true for all definitions of Sect. 3.4 So you ought to wear your boots in case you go out, as it should be.
- To refute the 'naïve approach,' we used the set of prioritized imperatives $p_1 < p_1 \Rightarrow p_2 < \neg p_2$. Since $<$ is already fully prioritized, the construction produces just one maximally obeyable subset, which is $\{!p_1,\ p_1 \Rightarrow !p_2\}$, as its two imperatives get added in the first two steps, and nothing is added in the third since $m(I)$ is inconsistent. All truth definitions make true $O(p_1/\top)$, none makes true the non-intuitive formula $O(p_1 \wedge \neg p_2/\top)$, and definitions that accept 'deontic detachment' make true $O(p_1 \wedge p_2/\top)$. So you must go out and wear a scarf, which is as it should be.
- To refute the stepwise approach the ordering of the imperatives was changed into $p_1 \Rightarrow p_2 < p_1 < \neg p_2$. Still $\mathscr{P}^{o}_{\mathcal{I}}(\top\}, I) = \{\{!p_1,\ p_1 \Rightarrow !p_2\}\}$, so the sentences made true by the truth definitions of Sect. 3.4 likewise do not change, and in particular the non-intuitive formula $O(p_1 \wedge \neg p_2/\top)$ is still false, and definitions that accept 'deontic detachment' make true $O(p_1 \wedge p_2/\top)$, so again you must go out and wear a scarf, which is as it should be.
- To refute the reconsidering and the fixpoint approaches the ordering of the imperatives was again changed into $p_1 \Rightarrow !p_2 < !\neg p_2 < !p_1$. Now $\mathscr{P}^{o}_{\mathcal{I}}(\top\}, I) = \{\{p_1 \Rightarrow !p_2, !\neg p_2\}\}$. All truth definitions make true $O(\neg p_2/\top)$ but not $O(p_1/\top)$ so the agent must satisfy the second ranking imperative, but not the third ranking imperative, which otherwise would include violating the highest ranking imperative, which is as it should be.
- Troublesome for the fixpoint approach was also the set of prioritized imperatives $p_1 \Rightarrow !p_2 < !(p_1 \wedge \neg p_2) < !p_3$: no fixpoint could be made out and so the approach produced no preferred subset, making everything obligatory. The preferred maximally obeyable subset is $\{p_1 \Rightarrow !p_2, !p_3\}$, eliminating the second ranking imperative that demands a violation of the first, and making $O(p_3/\top)$ true for all truth definitions, which again is as it should be.
- Finally, consider the 'drinking and driving' example: the set of prioritized imperatives $p_1 \Rightarrow !\neg p_2 < p_3 \Rightarrow !p_2 < p_3 \Rightarrow !p_1$ produces, for the situation p_3, the set of preferred maximally obeyable subsets $\mathscr{P}^{o}_{\mathcal{I}}(\{p_3\}, I) = \{\{p_1 \Rightarrow !\neg p_2, p_3 \Rightarrow !p_2\}\}$, making true $O(p_2/p_3)$ for all truth definitions of Sect. 3.4, so given that I go to the party I must do the driving, which is as it should be.

As could be seen, all truth definitions now produce the 'right' results in the examples used. Moreover, since all truth definitions refer to the same preferred subsets $\mathscr{P}^{o}_{\mathcal{I}}(\{C\}, I)$, it is possible to index the O-operators according to the truth definition employed, and e.g. state truths like $O^1(A/C) \wedge O^3(B/C) \rightarrow O^4(A \wedge B/C)$, meaning that if, for any maximal set of imperatives that I can obey in the situation C, imperatives are triggered that demand A, and that if I satisfy all such triggered imperatives, I will have to do B, then obeying a maximal number of imperatives includes having to do $A \wedge B$. It may well be that natural language 'ought-statements' are ambiguous in the face of conditional demands, the discussion in Sect. 3 suggested this. If maximal obeyability is accepted as the threshold criterion that limits what norms an agent can be expected to conform to in a given situation, then definition 8 leaves the philosophical logician with maximal freedom as to what deontic operator is chosen.

6 Theorems

Truth definitions (*td-pcd*$1^{(+)} - 4^{(+)}$) define when a sentence of the form $O(A/C)$ is true or false with respect to a prioritized conditional structure $\mathcal{I}$ and a situation C. So I briefly consider what sentences are theorems, i.e. hold for all such structures, given the usual truth definitions for Boolean operators. It is immediate that for all truth definitions, (DExt), (DM), (DC), (DN) and (DD-R) are theorems (cf. Sect. 3.1). (DD-R) states that there cannot be both an obligation to bring about A and one to bring about $\neg A$ unless the situation C is logically impossible, so our truth definitions succeed in eliminating conflicts. All these theorems are 'monadic' in the sense that C is kept fixed; in fact, they are the C-relative equivalents of standard deontic logic *SDL*. More interesting are theorems (known from the study of non-monotonic reasoning) that describe relations between obligations in different circumstances. Obviously we have

$$\text{(ExtC)}\quad \text{If } \vdash_{PL} C \leftrightarrow D \text{ then } O(A/C) \leftrightarrow O(A/D) \text{ is a theorem}$$

for all truth definitions, i.e. for equivalent situations C, the obligations do not change. As long as truth definitions are not sensitive to conflicts, e.g. for ($td - cd^{(+)} - 4^{(+)}$), we have 'strengthening of the antecedent,' i.e. for these definitions

$$\text{(SA)}\quad O(A/C) \to O(A/C \wedge D)$$

holds. When only maximally obeyable subsets are considered, i.e. for truth definitions (*td-pcd*$1^{(+)} - 4^{(+)}$), both (SA) and the weaker 'rational monotonicity' theorem

$$\text{(RM)}\quad \neg O(\neg D/C) \wedge O(A/C) \to O(A/C \wedge D)$$

are refuted e.g. by a set $I = \{!(p_1 \wedge p_2), !(p_1 \wedge \neg p_2), p_2 \Rightarrow \neg p_1\}$ of equally ranking imperatives: though $O(p_1/\top)$ is true and $O(\neg p_2/\top)$ false, $O(p_1/p_2)$ is false. However, for all definitions of Sect. 3.4, '(conjunctive) cautious monotonicity'

$$\text{(CCMon)}\quad O(A \wedge D/C) \to O(A/C \wedge D)$$

holds, which states that if you should to two things and you do one of them, you still have the other one left.[22] Moreover, truth definitions (*td-pcd*$1^+, 2^+, 3^+, 4^+$) validate the 'circumstantial extensionality' rule

$$\text{(CExt)}\quad \text{If } \vdash_{PL} C \to (A \leftrightarrow B) \text{ then } O(A/C) \leftrightarrow O(B/C) \text{ is a theorem}$$

that corresponds to 'circumstantial reasoning.' All definitions that accept 'reasoning by cases,' i.e. (*td-pcd*$2, 2^+, 4, 4^+$), make

$$(Or)\quad O(A/C) \wedge O(A/D) \to O(A/C \vee D)$$

a theorem. Note that (CExt) and (Or) derive

$$\text{(Cond)}\quad O(A/C \wedge D) \to O(D \to A/C),$$

which in turn derives (Or) in the presence of (DC), and that by adding (CExt) and (Or) we obtain again the system *PD* (cf. Sect. 3). Finally, all definitions with 'deontic detachment,' i.e. (*td-pcd*$3, 3^+, 4, 4^+$), make

$$\text{(Cut)}\quad O(A/C \wedge D) \wedge O(D/C) \to O(A/C)$$

[22] This is Hansson's [15] theorem (19).

a theorem. (Cut) is derivable given (Cond) (use Cond on the first conjunct $O(A/C \wedge D)$ to obtain $O(D \to A/C)$, agglomerate with $O(D/C)$, and from $O(D \wedge (D \to A)/C)$ derive $O(A/C)$), which syntactically mirrors the semantic equivalence of definitions (*td-pcd*2^+) and (*td-pcd*4^+). Theoremhood of all of the above for semantics that employ the respective truth definitions is easily proved and left to the reader (cf. Hensen [13] and [14] as well as Makinson and van der Torre [25] for the general outline). Makinson and van der Torre's results also suggest that these theorems axiomatically define complete systems of deontic logic with respect to semantics that employ the respective truth definitions (*td-pcd*$1^{(+)} - 4^{(+)}$), but this remains a conjecture that further study must corroborate.[23]

7 Back to the beginning: questions of representation

One might wonder if it is always adequate to represent a natural language conditional imperative 'if ... then bring about that ___' by use of a set I containing an imperative i with a $g(i)$ that formalizes '...' and a $f(i)$ that formalizes '___.' This is because there is a second possibility: represent the natural language conditional imperative by an unconditional imperative $!(g(i) \to f(i))$. We saw in Sect. 3 that this is not generally adequate. But that does not mean that such a representation is not *sometimes* what is required. Consider the crucial imperatives in the previous examples: perhaps what your mother meant was simply 'don't drink and drive;' perhaps what the doctor meant was 'don't go out without a scarf;' perhaps the Colonel meant to tell O'Reilly not to do both, turn the heat on and keep the window closed; perhaps self-defense required me to see to it that I am not attacked without fighting back. These interpretations seem not wholly unreasonable, and if they are adequate, then the best representation would be by an imperative $!(g(i) \to f(i))$ instead of $g(i) \Rightarrow !f(i)$.

What then are the conditions that make a representation by an unconditional imperative adequate? One test may be to ask: 'Would bringing about the absence of the antecedent condition count as satisfaction of the imperative?' Would not drinking, not going out, not turning on the heat, making the man not attack, count as properly reacting to the imperatives in question? It should be if what the imperatives demand is a material conditional, since then the conditional imperatives in question are equivalent to telling the agent 'either don't drink or don't drive, its your decision,' 'either don't go out, or wear a scarf,' 'either don't turn on the heat, or open the window,' etc. Another test would be to examine if contraposition is acceptable. Can we say that your mother wanted you not to drink if you are going to drive, that the doctor wanted you to stay inside if you are not going to wear a scarf, that the Colonel wanted O'Reilly to turn off the heat if the window is closed, that self-defense requires you to make the man not attack if you are not going to fight back? If the proper representation is by imperatives that demand a material conditional, then the answers should be affirmative. I do not think these are easy questions, however, and leave them to the reader to discuss and answer at his or her own discretion. But it is easy to see that, had we chosen to represent the crucial imperatives in the above examples (including the 'drinking and driving' problem) by unconditional imperatives that demand a material implication to be realized, then all of the discussed methods would have resolved these examples.

[23] For (*td-pcd*2^+, 4^+), completeness of *PD* is immediate from the results in [13], [14].

8 Conclusion

Reasoning about obligations when faced with different and possibly conflicting imperatives is a part of everyday life. To avoid conflicts, these imperatives may be ordered by priority and then observed according to their respective ranks. The ‘drinking and driving’ case in the introduction presented an example of such natural reasoning. Providing a formal account is, however, additionally complicated by the fact that there are various and mutually exclusive intuitions about what belongs to the right definition of an ‘obligation in the face of conditional imperatives,’ i.e. the definition of a deontic O-operator. Based on similar definitions of operators by Makinson and van der Torre [25], [26] for their ‘input/output logic’, but leaving the choice of the ‘right’ operator to the reader, I presented several proposals in Sect. 3 for definitions of a dyadic O-operator, namely (*td-pcd*$1^{(+)} - 4^{(+)}$). These were dependent on a choice of ‘preferred subsets’ among a given set of prioritized conditional imperatives. A particularly successful method to identify such subsets, but applying to unconditional imperatives only, was Brewka’s [3] definition of ‘preferred subtheories’ within a theory. In Sect. 4 I discussed various approaches that extend this method to conditional imperatives, but these failed to produce satisfactory results for a number of given examples. In Sect. 5 I proposed that the maximality criterion used to construct the preferred subsets should be the avoidance of conflict. A recent approach to this task by ‘consistent extensions’ was found to be biased towards definitions of obligation that do not accept reasoning by cases; another, that aims to avoid the truth of $O(\neg C/C)$ for possible circumstances C, produced satisfactory results only for truth definitions that accept deontic detachment. I then argued that the solution is to adapt Brewka’s method in such a way that it constructs, instead of maximal subsets of imperatives that are collectively satisfiable by an agent, maximally *obeyable* subsets of the imperatives, in the sense that the facts do not derive that some imperative of the set must be violated. I showed that this new proposal provides adequate solutions to all of the examples, and in particular the ‘drinking and driving’ example is resolved in a satisfactory fashion for all of the discussed deontic operators.

Acknowledgments I am grateful to Lou Goble, John F. Horty, David Makinson and Leon van der Torre for helpful comments and discussions. Also, I thank four anonymous referees for pointed remarks and useful suggestions. A preliminary version of this paper was presented to the participants in the seminar “Normative Multi-agent Systems (NorMAS’07)” held in Dagstuhl, Germany, from 18 to 23 March 2007.

References

1. Alchourrón, C. E., & Bulygin, E. (1981). The expressive conception of norms. In [18], 95–124.
2. Alchourrón, C. E., & Makinson, D. (1981). Hierarchies of regulations and their logic. In R. Hilpinen (Ed.), *New Studies in Deontic Logic* (pp. 125–148). Dordrecht: Reidel.
3. Brewka, G. (1989). Preferred subtheories: An extended logical framework for default reasoning. In N. S. Sridharan (Ed.), *Proceedings of the eleventh international joint conference on artificial intelligence IJCAI-89, Detroit, Michigan, USA, August 20–25, 1989* (pp. 1043–1048). San Mateo, Calif.: Kaufmann.
4. Brewka, G. (1994). Reasoning about priorities in default logic. In B. Hayes-Roth & R. E. Korf (Eds.), *Proceedings of the 12th national conference on artificial intelligence, Seattle, WA, July 31st–August 4th, 1994* (Vol. 2 pp. 940–945). Menlo Park: AAAI Press.
5. Brewka, G., & Eiter, T. (1999). Preferred answer sets for extended logic programs. *Artificial Intelligence, 109*, 297–356.
6. Broersen, J., Dastani, M., & van der Torre, L. (2002). Realistic desires. *Journal of Applied Non-Classical Logics, 12*, 287–308.
7. Caminada, M., & Sakama, C. (2006). On the existence of answer sets in normal extended logic programs. In G. Brewka, S. Coradeschi, A. Perini, & P. Traverso (Eds.), *Proceedings of the 17th European con-*

ference on artificial intelligence (ECAI 2006), August 29–September 1, 2006, Riva del Garda, Italy (pp. 741–742). Amsterdam: IOS Press.

8. Downing, P. (1959). Opposite conditionals and deontic logic. *Mind, 63*, 491–502.
9. van Fraassen, B. (1973). Values and the heart's command. *Journal of Philosophy, 70*, 5–19.
10. Goble, L. (2005). A logic for deontic dilemmas. *Journal of Applied Logic, 3*, 461–483.
11. Greenspan, P. (1975). Conditional oughts and hypothetical imperatives. *Journal of Philosophy, 72*, 259–276.
12. Hansen, J. (2004). Problems and results for logics about imperatives. *Journal of Applied Logic, 2*, 39–61.
13. Hansen, J. (2005). Conflicting imperatives and dyadic deontic logic. *Journal of Applied Logic, 3*, 484–511.
14. Hansen, J. (2006). Deontic logics for prioritized imperatives. *Artificial Intelligence and Law, 14*, 1–34.
15. Hansson, B. (1969). An analysis of some deontic logics. *Nôus, 3*, 373–398. Reprinted in [17], 121–147.
16. Hansson, S. O. (2001). *The structure of values and norms*. Cambridge: Cambridge University Press.
17. Hilpinen, R. (Ed.) (1971). *Deontic logic: Introductory and systematic readings*. Dordrecht: Reidel.
18. Hilpinen, R. (Ed.) (1981). *New studies in deontic logic*. Dordrecht: Reidel.
19. Hofstadter, A., & McKinsey, J. C. C. (1938). On the logic of imperatives. *Philosophy of Science, 6*, 446–457.
20. Horty, J. F. (2003). Reasoning with moral conflicts. *Noûs, 37*, 557–605.
21. Horty, J. F. (2007). Defaults with priorities. *Journal of Philosophical Logic, 36*, 367–413.
22. Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence, 44*, 167–207.
23. Lewis, D. (1974). Semantic analysis for dyadic deontic logics. In S. Stenlund (Ed.), *Logical theory and semantic analysis* (pp. 1–14). Dordrecht: Reidel.
24. Makinson, D. (1999). On a fundamental problem of deontic logic. In P. McNamara & H. Prakken (Eds.), *Norms, logics and information systems* (pp. 29–53). Amsterdam: IOS.
25. Makinson, D., & van der Torre, L. (2000). Input/output logics. *Journal of Philosophical Logic, 29*, 383–408.
26. Makinson, D., & van der Torre, L. (2001). Constraints for Input/output logics. *Journal of Philosophical Logic, 30*, 155–185.
27. Marek, V. W., & Truszczyński, M. (1993). *Nonmonotonic logic context-dependent reasoning*. Berlin: Springer.
28. Nebel, B. (1991). Belief revision and default reasoning: Syntax-based approaches. In J. A. Allen, R. Fikes & E. Sandewall (Eds.), *Principles of knowledge representation and reasoning: Proceedings of the second international conference, KR '91, Cambridge, MA, April 1991* (pp. 417–428). San Mateo: Morgan Kaufmann.
29. Nebel, B. (1992). Syntax-based approaches to belief revision. In: P. Gärdenfors (Ed.), *Belief revision* (pp. 52–88). Cambridge: Cambridge University Press.
30. Prakken, H. (1997). *Logical tools for modelling legal argument*. Dordrecht: Kluwer.
31. Prakken, H., & Sartor, G. (1997). Argument-based logic programming with defeasible priorities. *Journal of Applied Non-classical Logics, 7*, 25–75.
32. Rescher, N. (1964). *Hypothetical reasoning*. Amsterdam: North-Holland.
33. Rescher, N. (1966). *The Logic of commands*. London: Routledge & Kegan Paul.
34. Rintanen, J. (1994). Prioritized autoepistemic logic. In: C. MacNish, D. Pearce, & L. M. Pereira (Eds.), *Logics in artificial intelligence, European Workshop, JELIA '94, York, September 1994, Proceedings* (pp. 232–246). Berlin: Springer.
35. Ross, W. D. (1930). *The right and the good*. Oxford: Clarendon Press.
36. Ryan, M. (1992). Representing defaults as sentences with reduced priority. In B. Nebel, C. Rich, & W. Swartout (Eds.), *Principles of knowledge representation and reasoning: Proceedings of the third international conference, KR '92, Cambridge, MA, October 1992*. (pp. 649–660). San Mateo: Morgan: Kaufmann.
37. Sakama, C., & Inoue, K. (1996). Representing priorities in logic programs. In: M. Maher (Ed.), *Joint international conference and symposium on logic programming JICSLP 1996, Bonn, September 1996* (pp. 82–96). Cambridge: MIT Press.
38. Sosa, E. (1996). The logic of imperatives. *Theoria, 32*, 224–235.
39. Świrydowicz, K. (1994). Normative consequence relation and consequence operations on the language of dyadic deontic logic. *Theoria, 60*, 27–47.
40. von Wright, G. H. (1968). *An essay in deontic logic and the general theory of action*. Amsterdam: North Holland.
41. von Wright, G. H. (1984). Bedingungsnormen, ein Prüfstein für die Normenlogik. In W. Krawietz, H. Schelsky, O. Weinberger, & G. Winkler (Eds.), *Theorie der Normen* (pp. 447–456). Berlin: Duncker & Humblot.

42. Watts, I. (1725). *Logick: or, the right use of reason in the inquiry after truth, with a variety of rules to guard against error in the affairs of religion and human life, as well as in the sciences*. London: for John Clark, Richard Hett et al.