What is Input/Output Logic?

David Makinson [a] Leendert van der Torre [b]

[a] King's College London, Email: dcmakinson@aol.com
[b] Vrije Universiteit Amsterdam, Email: torre@cs.vu.nl

Abstract

We explain the *raison d'être* and basic ideas of input/output logic, sketching the central elements with pointers to other publications for detailed developments. The motivation comes from the logic of norms. Unconstrained input/output operations are straightforward to define, with relatively simple behaviour, but ignore the subtleties of contrary-to-duty norms. To deal with these more sensitively, we constrain input/output operations by means of consistency conditions, expressed via the concept of an outfamily. However, this is a more complex affair, with difficult choices between alternative options.

1 Motivation

Input/output logic takes its origin in the study of conditional norms. These may express desired features of a situation, obligations under some legal, moral or practical code, goals, contingency plans, advice, etc. Typically they may be expressed in terms like: *In such-and-such a situation, so-and-so should be the case*, or *...should be brought about*, or *...should be worked towards*, or *...should be followed* – these locutions corresponding roughly to the kinds of norm mentioned.

To be more accurate, input/output logic has its source in a tension between the philosophy of norms and formal work of deontic logicians.

Philosophically, it is widely accepted that a distinction may be drawn between norms on the one hand, and declarative statements on the other. Declarative statements may bear truth-values, in other words are capable of being true or false; but norms are items of another kind. They may be respected (or not), and may also be assessed from the standpoint of other norms, for example when a legal norm is judged from a moral point of view (or vice versa). But it makes no sense to describe norms as true or as false.

However the formal work of deontic logicians often goes on as if such a distinction had never been heard of. The usual presentations of deontic logic, whether axiomatic or semantic, treat norms as if they could bear truth-values. In particular, the truth-functional connectives *and*, *or* and most spectacularly *not* are routinely applied to norms, forming compound norms out of elementary ones. Semantic constructions using possible worlds go further by offering rules to determine, in a model, the truth-value of a norm.

This anomaly was noticed more than half a century ago, by Dubislav [4] and Jørgensen [5], but little was done about it. Indeed, from the 1960s onwards, the semantic approach in terms of possible worlds deepened the gap. The first serious attempt by a logician to face the problem appears to be due to Stenius [13], followed by Alchourrón and Bulygin [2] for unconditional norms, then Alchourrón [1] and Makinson [7] for conditional ones. Input/output logic may be seen as an attempt to extract the essential mathematical structure behind these reconstructions of deontic logic.

Like every other approach to deontic logic, input/output logic must face the problem of accounting adequately for the behaviour of what are called 'contrary-to-duty' norms. The problem may be stated thus: given a set of norms to be applied, how should we determine which obligations are operative in a situation that already violates some among them? It appears that input/output logic provides a convenient platform for dealing with this problem by imposing consistency constraints on the generation of output.

We begin by outlining the central ideas and constructions of unconstrained input/output logic. These are quite straightforward, and provide the basic framework of the theory. We then sketch a strategy for constraining those operations so as to deal more sensitively with contrary-to-duty situations. For further details, the reader is invited to refer to Makinson and van der Torre [8, 9].

2 Unconstrained Input/Output Operations

We avoid assuming that conditional norms bear truth-values. They are not embedded in compound formulae using truth-functional connectives. To avoid all confusion, they are not even treated as formulae, but simply as ordered pairs (a, x) of purely boolean (or eventually first-order) formulae.

Technically, a normative code is seen as a set G of conditional norms, i.e. a set of such ordered pairs (a, x). For each such pair, the body a is thought of as an *input*, representing some condition or situation, and the head x is thought of as an *output*, representing what the norm tells us to be desirable, obligatory or whatever in that situation. The task of logic is seen as a modest one. It is not to create or determine a distinguished set of norms, but rather to prepare information before it goes in as input to such a set G, to unpack output as it emerges and, if needed, coordinate the two in certain ways. A set G of conditional norms is thus seen as a transformation device, and the task of logic is to act as its 'secretarial assistant'.

The simplest kind of unconstrained input/output operation is depicted in Figure 1. A set A of propositions serves as explicit input, which is prepared by being expanded to its classical closure $Cn(A)$. This is then passed into the 'black box' or 'transformer' G, which delivers the corresponding immediate output $G(Cn(A)) = \{x \mid \text{ for some } a \in Cn(A), (a, x) \in G\}$. Finally, this is expanded by classical closure again into the full output $out_1(G, A) = Cn(G(Cn(A)))$. We call this *simple-minded output.*

This is already an interesting operation. As desired, it does not satisfy the principle of identity, which in this context we call *throughput*, i.e. in general we do

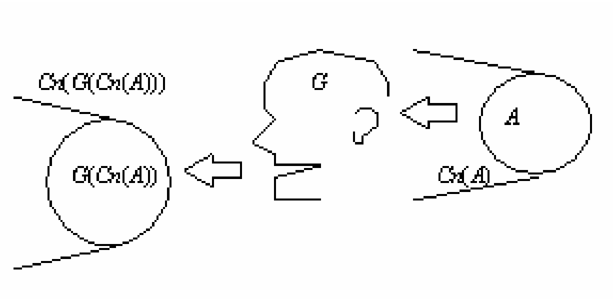

Figure 1: Simple-minded Output $out_1(G, A) = Cn(G(Cn(A)))$

not have $a \in out_1(G, \{a\})$ – which we write briefly, dropping the parentheses, as $out_1(G, a)$. It is characterized by three rules. Writing $x \in out_1(G, a)$ as $(a, x) \in out_1(G)$ and dropping the right hand side as G is held constant, these rules are:

Strengthening Input (SI):	From (a, x) to (b, x) whenever $a \in Cn(b)$
Conjoining Output (AND):	From (a, x), (a, y) to $(a, x \wedge y)$
Weakening Output (WO):	From (a, x) to (a, y) whenever $y \in Cn(x)$.

But simple-minded output lacks certain features that may be desirable in some contexts. In the first place, the preparation of inputs is not very sophisticated. Consider two inputs a and b. By classical logic, if $x \in Cn(a)$ and $x \in Cn(b)$ then $x \in Cn(a \vee b)$. But there is nothing to tell us that if $x \in out_1(G, a) = Cn(G(Cn(a)))$ and $x \in out_1(G, b) = Cn(G(Cn(b)))$ then $x \in out_1(G, a \vee b) = Cn(G(Cn(a \vee b)))$.

In the second place, even when we do not want inputs to be automatically carried through as outputs, we may still want outputs to be reusable as inputs – which is quite a different matter.

Operations satisfying each of these two features can be provided with explicit definitions, pictured by diagrams in the same spirit as that for simple-minded output, and characterized by straightforward rules. We thus have four very natural systems of input/output, which are labelled as follows: *simple-minded* alias out_1 (as above), *basic* (simple-minded plus input disjunction: out_2), *reusable* (simple-minded plus reusability: out_3), and *reusable basic* (all together: out_4).

For example, reusable basic output may be given a diagram and definition as in Figure 2. In the definition, a complete set is one that is either maximally consistent or equal to the set of all formulae.

The three stronger systems may also be characterized by adding one or both of the following rules to those for simple-minded output:

Disjoining input (OR):	From (a, x), (b, x) to $(a \vee b, x)$
Cumulative transitivity (CT):	From (a, x), $(a \wedge x, y)$ to (a, y).

These four operations have four counterparts that also allow *throughput*. Intu-

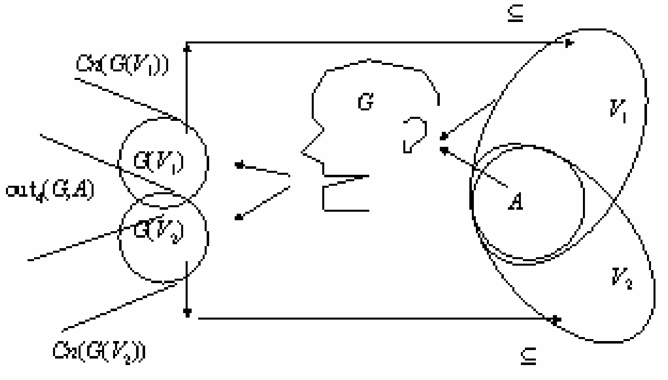

Figure 2: Basic Reusable Output $out_4(G,A) = \cap\{Cn(G(V)) \mid A \subseteq V \supseteq G(V), V \text{ complete}\}$

itively, this amounts to requiring $A \subseteq G(A)$. In terms of the definitions, it is to require that G is expanded to contain the diagonal, i.e. all pairs (a,a). Diagrammatically it is to add arrows from G's ear to mouth. Derivationally, it is to allow arbitrary pairs of the form (a,a) to appear as leaves of a derivation; this is called the zero-premise identity rule ID.

All eight systems are distinct, with one exception: basic throughput, which we write as out_2^+, authorizes reusability, so that $out_2^+ = out_4^+$. This may be shown directly in terms of the definitions, or using the following simple derivation of CT from the other rules.

$$\dfrac{\dfrac{\dfrac{\dfrac{(a,x)}{(a\wedge\neg x, x)}\,\text{SI} \quad \dfrac{-}{(a\wedge\neg x, a\wedge\neg x)}\,\text{ID}}{(a\wedge\neg x, x\wedge(a\wedge\neg x))}\,\text{AND}}{(a\wedge\neg x, y)}\,\text{WO} \qquad (a\wedge x, y)}{(a,y)}\,\text{OR}$$

The application of WO here is justified by the fact that $y \in Cn(x\wedge(a\wedge\neg x))$ since the right hand formula is a contradiction. Note that all rules available in basic throughput (including, in particular, identity) are needed in the derivation, reflecting the fact that CT is not derivable in the weaker systems.

This strong system indeed collapses into classical consequence, in the sense that $out_4^+(G,A) = Cn(m(G)\cup A)$ where $m(G)$ is the materialization of G, i.e. the set of all formulae $a \to x$ where $(a,x) \in G$.

The authors' papers [8] and [9, section 1] investigate these systems in detail – semantically, in terms of their explicit definitions, derivationally, in terms of the rules determining them, both separately and in relation to each other. We do not attempt to summarize the results here, but hope that the reader is tempted to follow further.

3 Why constrain?

As mentioned in section 1, all approaches to deontic logic must face the problem of dealing with contrary-to-duty norms. In general terms, we recall, the problem is: given a set of norms, how should we determine which obligations are operative in a situation that already violates some among them.

The following simple example is adapted from Prakken and Sergot [11].[1] Suppose we have the following two norms: *The cottage should not have a fence or a dog; if it has a dog it must have both a fence and a warning sign.*

In the usual deontic notation, where t stands for a tautology: $O(\neg(f \vee d)/t)$, $O(f \wedge w/d)$; in the notation of input/output logic: $(t, \neg(f \vee d))$, $(d, f \wedge w)$. Suppose further that we are in the situation that the cottage has a dog, thus violating the first norm. What are our current obligations?

Unrestricted input/output logic gives f: *the cottage has a fence* and w: *the cottage has a warning sign.* Less convincingly, because unhelpful if the presence of a dog is regarded as unalterable, it also gives $\neg d$: *the cottage does not have a dog.* Even less convincingly, it gives $\neg f$: *the cottage does not have a fence*, which is the opposite of what we want.

These results hold even for simple-minded output, without reusability or disjunction of inputs. The only rules needed are SI and WO, as shown by the following derivation of $\neg f$.

$$\cfrac{\cfrac{(t, \neg(f \vee d))}{(t, \neg f)}\ \text{WO}}{(d, \neg f)}\ \text{SI}$$

A common reaction to examples such as these is to ask: why not just drop the rule SI of strengthening the input? In semantic terms, why not cut back the definition of simple-minded output from $Cn(G(Cn(A)))$ to $Cn(G(A))$, and in similar (but more complex) fashion with the others? Indeed, this is a possible option, and the strategy that we will describe below does have the effect of disallowing certain applications of SI. But simply to drop SI is, in the view of the authors, too heavy-handed. We need to know *why* SI is not always appropriate and, especially, *when* it remains justified.

4 A Strategy for Constraint: Maxfamilies and their Outfamilies

Our strategy is to adapt a technique that is well known in the logic of belief change – cut back the set of norms to just below the threshold of making the

[1] There are many examples in the literature. Most of them involve ingredients that, while perfectly natural in ordinary discourse, are extraneous to the essential problem and thus invite false analyses. These ingredients include defeasibility, causality, the passage of time, and the use of questionable rules such as CT and OR in deriving output. We have chosen a very simple example that avoids all those elements. There is one respect in which it could perhaps be further purified: under input d, the output is not only inconsistent with the input, but also itself inconsistent. This matter is discussed at the end of section 5.

current situation contrary-to-duty. In effect, we carry out a contraction on the set G of given norms.

Specifically, we look at the maximal subsets $G' \subseteq G$ such that $out(G', A)$ is consistent with input A. In Makinson and van der Torre (2001), the family of such G' is called the *maxfamily* of (G, A), and the family of outputs $out(G', A)$ for G' in the maxfamily, is called the *outfamily* of (G, A).[2]

To illustrate this, consider the cottage example where $G = \{(t, \neg(f \vee d)), (d, f \wedge w)\}$, with the contrary-to-duty input d. Using simple-minded output, $maxfamily(G, d)$ has just one element $\{(d, f \wedge w)\}$, and so $outfamily(G, d)$ has one element, namely $Cn(f \wedge w)$.

Although the outfamily strategy is designed to deal with contrary-to-duty norms, its application turns out to be closely related to belief revision and non-monotonic reasoning when the underlying input/output operation authorizes throughput.

When all elements of G are of the form (t, x), then for the degenerate input/output operation $out_2^+(G, a) = out_4^+(G, a) = Cn(m(G) \cup \{a\})$, the elements of $outfamily(G, a)$ are just the maxichoice revisions of $m(G)$ by a, in the sense of Alchourrón, Gärdenfors and Makinson [3]. These coincide, in turn, with the extensions of the default system $(m(G), a, \emptyset)$ of Poole [10].

More surprisingly, there are close connections with the default logic of Reiter, falling a little short of identity. Read elements (a, x) of G as normal default rules $a; x/x$ in the sense of Reiter [12], and write $extfamily(G, A)$ for the set of extensions of (G, A). Then, for reusable simple-minded throughput out_3^+, it can be shown that $extfamily(G, A) \subseteq outfamily(G, A)$ and indeed that $extfamily(G, A)$ consists of precisely the maximal elements (under set inclusion) of $outfamily(G, A)$.

These results and related ones are proven in Makinson and van der Torre [9]. But in accord with the motivation from the logic of norms, the main focus in that paper is on input/output logics *without* throughput. Two kinds of question are investigated in detail there.

4.1 The search for truth-functional reductions of the consistency constraint

From the point of view of computation, it is convenient to make consistency checks as simple as possible, and executable using no more than already existing programs. For this reason, it is of interest to ask: under what conditions is the

[2]So defined, the outfamily is not in general the same as the family of all maximal values of $out(G', A)$ consistent with A, for G' ranging over subsets of G. Every maximal value of $out(G', A)$ is in the outfamily, but not always conversely. For certain of our output operations, the two families do coincide, but not for others.

This can be shown by simple examples, such as the Möbius strip of Makinson [6, 7]. Put $G = \{(a, x), (x, y), (y, \neg a)\}$. Then, for $out = out_3$ or $out = out_4$, $maxfamily(G, a)$ has three elements, namely the three two-element subsets of G. As a result, $outfamily(G, a)$ also has three elements – $Cn(\emptyset)$, $Cn(x)$, and $Cn(\{x, y\})$. Of these, only the last is a maximal value of $out(G', A)$ consistent with A for G' ranging over subsets of G.

We add that in this example, not even $Cn(\{x, y\})$ is a maximal subset of $out(G, a)$ that is consistent with a, for clearly $Cn(\{x, y\}) \subset Cn(\{x, y, \neg a \vee z\}) \subset out(G, a)$. Care is thus needed to avoid confusing maxfamilies with related maximal sets.

consistency of A with $out(G, A)$ reducible to the consistency of A with the materialization $m(G)$ of G, i.e. with the set of all formulae $a \to x$ where $(a, x) \in G$?

It is easy to check that the latter consistency implies the former for all seven of our input/output operations. It turns out that we have equivalence for just two of them (reusable basic with and without identity).

On the level of derivations, the question can take a rather different form, with different answers. Given a derivation of (a, x) with leaves L, under what conditions is the consistency of a with $out(L, a)$ equivalent to its consistency with $m(L)$? Curiously, this holds for a wider selection of our input/output operations – in fact, for all of them except basic output. Even more surprisingly, for some of the operations (those without OR), the same reduction also holds with respect to the set $h(L)$ of heads x, and the set $f(L)$ of fulfilments $a \wedge x$, of elements (a, x) of L.

From this result on derivations, we can go back and sharpen the semantic one. When G is a *minimal* set with $x \in out(G, a)$ then, for each of our input/output operations other than basic output, a is consistent with $out(G, a)$ iff it is consistent with $m(G)$ – and for the operations without OR, with $h(G)$, $f(G)$.

4.2 More severe applications of the consistency check

From a practical point of view, whenever we constrain an operation to avoid excess production, the question arises: how cautious (timid) or brave (foolhardy) do we want to be? For input/output operations, this issue arises in different ways on the semantic and derivational levels. On the semantic level, once we have formed an outfamily we may ask: should we intersect, join, or choose from its elements to obtain a unique restrained output? On the level of derivations, it is natural to ask: do we want to apply the consistency check only at the root of a derivation, or at every step within it?

The policy of checking only at the root corresponds to the option, on the semantic level, of forming the join of the outfamily; while the stricter policy of checking at every step is an essentially derivational requirement. But whichever of the two we choose, it is of interest to know under what conditions they coincide. In other words, given a derivation of (a, x) with leaves L such that a is consistent with $out(L, a)$, under what conditions does it follow that for every node (b, y) in the derivation, b is consistent with $out(L, b)$? It turns out that for certain of the seven input/output operations (again, those without the OR rule) this result holds. For operations with OR but without the rule CT, a rather subtler result may be obtained.

One lesson of these rather intricate investigations is that the behaviour of the consistency constraint depends very much on the choice of input/output operation; in particular, the presence of the rule OR destroys some properties. Another lesson is that questions can take different forms, with different answers, on the semantic and derivational levels. Thirdly, a detour through derivations can sometimes sharpen semantic results.

5 Conclusions

Drawing together the threads of this paper, we emphasize the main points.

- Input/output logic seeks to extract the essential mathematical structure behind recent attempts to reconstruct deontic logic that avoid treating norms as if they had truth-values.

- Unconstrained input/output provides us with a simple and elegant construction with straightforward behaviour, but whose application to norms totally ignores the subtleties of contrary-to-duty obligations.

- On the other hand, output constrained using the outfamily strategy provides a way of dealing with contrary-to-duty obligations. Its behaviour is quite subtle, and depends considerably on the choice of background input/output operation, in particular on whether or not it authorizes the rule of disjunction of inputs.

References

[1] Alchourrón, Carlos (1993). Philosophical foundations of deontic logic and the logic of defeasible conditionals. In Meyer, J.J. and R.J. Wieringa (eds) *Deontic Logic in Computer Science*, 43-84. Wiley: New York.

[2] Alchourrón, Carlos and Eugenio Bulygin (1981). The expressive conception of norms. In Hilpinen, R. (ed.) *New Essays in Deontic Logic*, 95-124. Reidel: Dordrecht.

[3] Alchourrón, Carlos, Peter Gärdenfors and David Makinson (1985). On the logic of theory change: partial meet contraction and revision functions. *The Journal of Symbolic Logic* 50 510-530.

[4] Dubislav, Walter (1937). Zur Unbegründbarkeit der Forderungsätze. *Theoria* 3: 330-342.

[5] Jørgensen, Jørgen (1937-8). Imperatives and logic. *Erkenntnis* 7: 288-296.

[6] Makinson, David (1994). General Patterns in Nonmonotonic Reasoning. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, ed. Gabbay, Hogger and Robinson, Oxford University Press, pages 35-110.

[7] Makinson, David (1999). On a fundamental problem of deontic logic, In Paul McNamara and Henry Prakken eds *Norms, Logics and Information Systems. New Studies in Deontic Logic and Computer Science*, Amsterdam: IOS Press, Series: Frontiers in Artificial Intelligence and Applications, Volume 49, pp. 29-53.

[8] Makinson, David and Leendert van der Torre (2000). Input/output logics, *J. Philosophical Logic* 29: 383-408.

[9] Makinson, David and Leendert van der Torre (2001). Constraints for input/output logics, *J. Philosophical Logic* 30(2):155-185.

[10] Poole, David (1988). A logical framework for default reasoning, *Artificial Intelligence* 36: 27-47.

[11] Prakken, Henry and Marek Sergot. Contrary-to-duty obligations (1996). *Studia Logica* 57: 91-115.

[12] Reiter, Ray (1980). A logic for default reasoning, *Artificial Intelligence* 13: 81-132.

[13] Stenius, Erik (1963). Principles of a logic of normative systems. *Acta Philosophica Fennica* 16: 247-260.