TOWARD A FRAMEWORK FOR AGENCY, INEVITABILITY, PRAISE AND BLAME

PAUL McNAMARA

There is little work of a systematic nature in ethical theory or deontic logic on aretaic notions such as praiseworthiness and blameworthiness, despite their centrality to common-sense morality. Without more work, there is little hope of filling the even larger gap of attempting to develop frameworks integrating such aretaic concepts with deontic concepts of common-sense morality, such as what is obligatory, permissible, impermissible, or supererogatory. It is also clear in the case of aretaic concepts that agency is central to such appraisal, so some agential notions must be integrated with aretaic concepts as well. The current paper takes the first step in a larger project aimed at the closure of these gaps. Here I sketch a simple framework for the aretaic appraisal of an agent's performance, layered on top of a simple framework for agency and predetermination. In Part I, I develop the framework for agency, ability, and inevitability, combining elements of work by Brown, Elgesem, Carmo, Santos, and Jones. In Part II, drawing on work by Chisholm and Sosa on intrinsic preferability, I sketch and explore a framework for defining aretaic superiority, praiseworthiness, blameworthiness, neutrality, and indifference, etc., retaining proper links to agency.

Part I: Agency, Ability, and Inevitability

We will imagine that our background theory involves some agent, Jane Doe, existing at various worlds. In these worlds, this agent often exhibits her agency by bringing certain things about. Presumably, she does so by taking certain actions that result in certain propositions being true, the ones she has brought about. We also imagine that some things are predetermined for our agent in a given world, in the sense of being inescapable for her *as of now*: some things are now such that no exercise of her ability is consistent with those things not being the case. The "as of now" stress is to steer the reader away from thinking that the predetermined things are necessarily things fixed for all time. No fatalistic notion is implied. With this in mind, we introduce two primitive operators, an agency operator and a predetermination operator:

BAp: It is (now) Brought About (by our agent) that p

PRp: It is (as of now) Predetermined (for our agent) that p

Nordic Journal of Philosophical Logic, Vol. 5, No. 2, pp. 135–160.

We will model agency using minimal models. This allows for a sufficiently fine-grained approach to block the validation of absurdities like agents bringing about logical truths whenever they act. We introduce an *agency function* mapping worlds to sets of propositions (world-theoretically construed):

$$BA\text{: } W \rightarrow \text{Pow}(\text{Pow}(W))\text{, that is, } BA_i \subseteq \text{Pow}(W)).$$

BA_i then denotes the set of propositions (possible empty) that the agent brings about at i. It will then be true at i that *our agent Brings it About that p* if and only if the proposition expressed by p is in the set of propositions our agent brings about at i:

$$M \models_i \text{BAp iff } \|p\|^M \in BA_i$$

It will be convenient later on to have introduced the following shorthand:

$$BAX = \{j\text{: } X \in BAj\}\text{, where } X \subseteq W.$$

Call any such proposition a *performance*: a proposition asserting that our agent brings X about, for some X.

We introduce some definitions:

$$\text{ROp} =_{df} \text{BA}\neg\text{p}$$

$$\text{NRp} =_{df} \neg\text{BA}\neg\text{p}$$

$$\text{LOp} =_{df} \neg\text{BAp} \ \& \ \neg\text{BA}\neg\text{p}$$

$$\text{RFp} =_{df} \text{BA}\neg\text{BAp}$$

The first says that *it is Ruled Out by what our agent does that p* if and only if our agent brings it about that ¬p. Note that this notion does not apply to all things that are ruled out per se, but only to those that are specifically ruled out *by* our agent's exercise of her agency. So contradictions, the negations of laws of nature and of past events are not ruled out *by* what our agent now does. The second says *it is Not Ruled out by anything our agent does that p* if and only if our agent does not bring it about that ¬p. Laws of logic (which are necessarily ruled in) as well as contradictions (which are necessarily ruled out) are not things that are ruled out *by* our agent. The third says *our agent Leaves it Open that p* (does nothing that determines the status of p) if and only if our agent neither brings about p nor rules p out by what she does. Again, it does not follow from the fact that our agent leaves something open that it is open per se. LOp is consistent with its being fixed that p and consistent with its being fixed that $\neg$p, as long as it is not fixed *by* anything our agent has done. So, these notions are all intended to have a strong agential reading. The final notion is a familiar abbreviation for refraining, in quasi-English, *it is a case of Refraining by our agent that p* if and only if our agent sees to it that she does not bring it about that p. No agent brings about logical truths, but neither does an agent bring it about that she doesn't. It has nothing to do with what she does. So refraining from p is not the same thing as merely not bringing about p.

The derived truth conditions for our defined operators are:

$$\mathcal{M}\models_i \text{ROp iff W-}\|p\|^{\mathcal{M}} \in BA_i$$

$$\mathcal{M}\models_i \text{NRp iff W-}\|p\|^{\mathcal{M}} \notin BA_i$$

$$\mathcal{M}\models_i \text{LOp iff } \|p\|^{\mathcal{M}} \notin BA_i \ \& \ \text{W-}\|p\|^{\mathcal{M}} \notin BA_i$$

$$\mathcal{M}\models_i \text{RFp iff W-}BA\|p\|^{\text{M}} \in BA_i.$$

We will model predetermination using standard Kripke models, introducing an *ability-relation*, *CO*, relating one world to a second when what happens at the second is consistent with our agent's abilities in the first world:

$$CO \subseteq \text{W} \times \text{W}.$$

The worlds consistent with our agent's abilities at a given world, i, might then be thought of as the *i-accessible worlds*:

$$CO^{\text{i}} = \{\text{j}: \langle \text{i}, \text{j} \rangle \in CO\}.$$

It will also prove convenient to introduce a notation for *the set of all propositions consistent with our agent's abilities*:

$$CO_{\text{i}} = \{\text{X}: \text{X} \cap CO^{\text{i}} \neq \varnothing\}$$

CO^{i} contains every world consistent with our agent's abilities at i, whereas CO_{i} contains the set of propositions true at any such world.

Let me note here that the fact that there is a world *consistent with* my abilities where p holds is not intended to imply that p *is within* my abilities: since tautologies are true at all worlds, they will be true at all worlds consistent with my abilities, but they are not within any agent's abilities at any world.

We can then say that *it is Predetermined for our agent that p* at i if and only if every world consistent with our agent's abilities at i is a world where p is true:

$$M \models_i \text{PR}p \text{ iff } \forall j(COij \rightarrow \models_j p)$$

We introduce some definitions:

$$\text{ES}p =_{df} \neg \text{PR}p$$

$$\text{CO}p =_{df} \neg \text{PR} \neg p$$

$$\text{CL}p =_{df} \text{PR} \neg p:$$

$$\text{ID}p =_{df} \neg \text{PR}p \ \& \ \neg \text{PR} \neg p$$

It is ES*capable for our agent that p* if and only if it is not predetermined that p. *It is* CO*nsistent with our agent's abilities that p* if and only if it is not predetermined for our agent that ¬p; *It is* CL*osed for our agent that p* if and only if it is predetermined that ¬p. *It is Indeterminate for our agent that p* if and only if it is neither predetermined for our agent that p nor predetermined for our agent that ¬p.

The derived truth conditions for these operators are:

$$M \models_i \text{ES}p \text{ iff } \exists j(COij \ \& \models_j \neg p)$$

$$M \models_i \text{CO}p \text{ iff } \exists j(COij \ \& \models_j p)$$

$$M \models_i \text{CL}p \text{ iff } \forall j(COij \rightarrow \models_j \neg p)$$

$$M \models_i \text{ID}p \text{ iff } \exists j(COij \ \& \models_j \neg p) \ \& \ \exists j(COij \ \& \models_j p)$$

We will endorse the following plausible constraints on BA_i and *CO*. For any world, i and proposition, X,

$$BA\text{-(t)}: X \in BA_i \rightarrow i \in X$$

$$BA\text{-(c)}: (X \in BA_i \ \& \ Y \in BA_i) \rightarrow (X \cap Y) \in BA_i$$

$$BA\text{-NO}: W \notin BA_i$$

$$CO\text{-RFLX}: COii$$

BA-(t) says that the propositions brought about by our agent at i are true at i. *BA*-(c) says that the conjunction of any two propositions among those now brought about by our agent at i is also now brought about by her at i. *BA*-NO says that the necessary proposition is not among those brought about by our agent. And *CO*-RFLX says that any world is consistent with the abilities our agent has at that world.

The following basic principles are validated:

BA-T:	$\models BAp \rightarrow p$
BA-C:	$\models (BAp \ \& \ BAq) \rightarrow BA(p \ \& \ q)$
BA-NO:	$\models \neg BA\mathit{t}$
PR-K:	$\models PR(p \rightarrow q) \rightarrow (PRp \rightarrow PRq)$
PR-T:	$\models PRp \rightarrow p$
BA-RE:	If $\models p \leftrightarrow q$ then $\models BAp \leftrightarrow BAq$
PR-NEC:	If $\models p$ then $\models PRp$

Using a Chellas-style naming convention, the pure BA component, TEC-NO, is the classical base logic for agency in Santos and Carmo 1996 and Santos, Jones and Carmo 1997, drawn from Elgesem 1993. The pure PR logic is the normal modal logic KT used for predetermination and kin in Brown 1992. Call the system consisting of the above principles, TEC-NO-KT.

The following are derivable in TEC-NO-KT:

BA-OD:	$\vdash \neg BA\mathit{f}$
BACO:	$\vdash BAp \rightarrow COp$
BA-NC:	$\vdash BAp \rightarrow \neg BA \neg p$
OD-NR:	$\vdash NR\mathit{f}$
NO-NR:	$\vdash NR\mathit{t}$
OD-LO:	$\vdash LO\mathit{f}$
NO-LO:	$\vdash LO\mathit{t}$
NR-T:	$\vdash p \rightarrow NRp$
CO-T:	$\vdash p \rightarrow COp$
N-CO:	$\vdash CO\mathit{t}$
	$\vdash BAp \rightarrow BA(p \ \& \ \mathit{t})$
	$\vdash BA(p \ \& \ \mathit{t}) \rightarrow (BAp \ \& \ \neg BA\mathit{t})$
NEC-CO:	If $\vdash p$ then $\vdash COp$
	If $\vdash p \rightarrow q$ then $\vdash BAp \leftrightarrow BA(p \ \& \ q)$
NR-RE:	If $\vdash p \leftrightarrow q$ then $\vdash NRp \leftrightarrow NRq$
	If $\vdash q \leftrightarrow \mathit{t}$ then $\vdash BA(p \ \& \ q) \rightarrow (BAp \ \& \ \neg BAq)$

We now follow a strategy in Brown 1992 for defining ability (except that our underlying logic for agency deviates from Brown 1992 in ways that make it more plausible as a logic for agency, and derivatively as a logic for ability):

$$\mathrm{ABp} =_{\mathrm{df}} \mathrm{COBAp}$$

So *it is within our agent's* AB*ility that p* if and only if it is consistent with our agent's abilities that she brings about p. Consequently, ABp is true at i if and only if there is a world consistent with our agent's i-based abilities where our agent brings it about that p:

$$M \models_{\mathrm{i}} \mathrm{ABp} \text{ iff } \exists \mathrm{j}(CO\mathrm{ij} \ \& \ M \models_{\mathrm{j}} \mathrm{BAp})$$

It will be convenient later on to have the following shorthands available:

$$AB_{\mathrm{i}} = \{\mathrm{X}: BA\mathrm{X} \in CO_{\mathrm{i}}\}$$
$$AB\mathrm{X} = \{\mathrm{j}: \mathrm{X} \in AB_{\mathrm{j}}\}.$$

AB_{i} is the set of all propositions within our agents abilities at i, and $AB\mathrm{X}$ is just the proposition asserting that X is within our agent's ability.

We said earlier that CO is intended to mean that *it is consistent with* our agents ability that p, not that *it is within* our agent's ability that p, and this analysis accords with that intuition. It expresses the idea that a) if something is within my ability, then it must be consistent with my abilities that I bring it about, and b) if it is consistent with my abilities that I bring something about, then it must be within my abilities to bring that thing about. The first seems plainly right, but the second is less clear. We can read "CO^{i}", the set of worlds consistent with our agent's abilities in i, two ways. In the first way, the worlds in CO^{i} must be consistent with the abilities *and the disabilities* that I have in i, so that in all these worlds, I have precisely the same abilities and disabilities that I have in i. Read this way, b) is plausible. Read a second, weaker way, the worlds in CO^{i} must be consistent with the abilities, but not necessarily the disabilities, that I have in i. Construed this way, in some of these worlds, although I will have all the abilities there that I have in i, I will have additional abilities as well. Furthermore, surely in some of these expanded-ability accessible worlds, I will be exercising some ability I have there but lack here. So on this reading there will be worlds in CO^{i} where I am exercising abilities I lack at i, and thus I will be bringing about things there that are not within my ability at i. Thus on this second reading, the analysis is not plausible at all. For on this reading, there will be things I bring about consistent with my abilities at i that are nonetheless not *within* my ability *at i*. So, given the intended interpretation, CO^i must be constrained to be the set of worlds consistent with both the abilities *and disabilities* that our agent has at i.

This issue of reading *CO* so that it does not relate one world to other worlds where my abilities are a genuine extension of my abilities in the first world is closely related to issues connected with iterated operators, and Brown 1992 is fascinating and instructive in its exploration of the extent to which agency and ability can be represented by two normal modal operators, with stress on iteration for the accessibility relation used for BA. However, Brown (1992) doesn't note that the analysis of ability is only plausible if CO^i is constrained in ways that rule out expanded-ability accessibility. Also matters turn out to involve certain subtleties for the combination herein of a classical logic for BA with a normal modal logic for PR that aren't found in more familiar combinations of two normal modal logics. We have seen some results already, but things become especially interesting for iteration of operators, particularly for BA. For example, among other perhaps surprising results, adding BA-B, $p \rightarrow BANRp$, or BA-5, $NRp \rightarrow BANRp$, to the system TEC-NO-KT results in outright inconsistency. So iteration turns out to be subtler here than it is with bimodal normal modal logics. Let me also briefly note that aside from TEC-NO-KT's appearing to blend some of the best elements of Brown 1992 and Santos and Carmo 1996, one central semantic notion in Elgesem 1993 and 1997 appears to be definable using *CO* and *BA*. Elgesem uses a function, *f*, from a world-proposition pair to a set of worlds: $f(i, X)$ is intended to denote a set of worlds (possible empty) where the agent realizes the ability (possibly vacuous) she possess in i to bring about X. In our framework, this appears to be definable as follows:

$$f(i, X) = \{j: j \in BAX \ \& \ X \in AB_i\}$$

Note that $f(i, X)$ need not be a set of worlds consistent with our agent's abilities in i, even on our weaker reading, for, in j, I can realize my non-vacuous ability in i to bring about X while lacking other i-based abilities in j. (Elgesem's groundbreaking work analyses an impressively wide array of agency-related notions with philosophical sensitivity, employing a second semantic notion that can't be represented with our resources.) A full exploration of the framework endorsed for agency here, and in particular of constraints associated with iteration for both the BA-based operators and the PR-based operators is beyond the scope of this paper, and will be taken up elsewhere where agency is in central focus.

We list below some additional principles that involve the operators associated with agency and predetermination, including principles for

AB, which is defined via such a mix:

DEF AB:	$\vdash \text{ABp} \leftrightarrow \neg \text{PR} \neg \text{BAp}$
AB-NO:	$\vdash \neg \text{AB}t$
AB-OD:	$\vdash \neg \text{AB}f$
BA-AB:	$\vdash \text{BAp} \rightarrow \text{ABp}$
ABBA-AB:	$\vdash \text{ABBAp} \rightarrow \text{ABp}$
	$\vdash \text{ABROp} \rightarrow \text{AB(ROp} \ \& \ \neg \text{BAp)}$
ABBA-NO:	$\vdash \neg \text{ABBA}t$
ABBA-OD:	$\vdash \neg \text{ABBA}f.$
	If $\vdash$ p $\rightarrow$ q then $\vdash$ ABp$\leftrightarrow$AB(p & q)
AB-RE:	If $\vdash$ p$\leftrightarrow$q then $\vdash$ ABp$\leftrightarrow$ABq

Carmo and Santos 1996 entertains, but does not endorse, the possible intuitive soundness, of a weakened version of RM, alluding to something like the following CO-*qualified version of RM*:

COQ-RM: If $\vdash$p $\rightarrow$ q then $\vdash$COq $\rightarrow$ (BAp $\rightarrow$ BAq)

However, I think this weakening is problematic on reflection. Let p be *I call you today*, and let q be *someone calls you today*. Now suppose that I will call you today, but cannot access a phone until late afternoon. Now add that it is inescapable for me that another person will call you this morning. In that case, this other person will inescapably bring q about, so that q will be settled, without my agency, before I ever bring about p. So q is not even within my abilities, even though p entails q, I do bring about p, and COq holds (since q does). More technically, since q could be *t*, any p entails *t*, CO*t* is an axiom, as is ¬BAt, the above rule would entail ¬BAp, for any p, eradicating all agency. So, we must reject the principle.

It might be thought that an AB-*qualified version of RM* might do:

ABQ-RM: If $\vdash$p $\rightarrow$ q then $\vdash$ABq $\rightarrow$ (BAp $\rightarrow$ BAq)

Here we replace "COp" with the stronger "ABp". This does survive the above counterexamples, but not for long. Suppose things are as before except that we now add that I could call you at any time of the day, including earlier than any other person, so q is within my ability, but *only if* I exercise my ability to bring about p early in the day. Now just add that I in fact don't, and call later instead, after the other person.

In each counterexample, the problem is that q need not reflect *my* agency. This might suggest qualifying RM to those cases where q must *reflect my agency*. This notion of an *agentially reflective* proposition is independently important anyway:

$$\text{ARp} =_{\text{df}} \text{COp} \;\&\; \text{PR}(\text{p} \rightarrow \text{BAp}).$$

So *it is Agentially Reflective that p* if and only if p is consistent with my abilities and it is predetermined that: p is true only if I bring it about that p. With this, we can state an AR-*Qualified version of RM*:

$$\text{ARQ-RM: If } \vdash \text{p} \rightarrow \text{q then } \vdash \text{ARq} \rightarrow (\text{BAp} \rightarrow \text{BAq}).$$

The fact that this is intuitively plausible is a happy one, since it is derivable: For suppose $\vdash$p $\rightarrow$ q. Now assume ARq and BAp. Since BAp, p by BA-T, and then since $\vdash$p $\rightarrow$ q, we get q. Now since ARq, PR(q $\rightarrow$ BAq). But then from PR-T, q $\rightarrow$ BAq. So BAq.

These are also derivable:

$$\vdash \neg \text{AR}t$$

$$\vdash \neg \text{AR}f$$

$$\vdash \text{ARp} \leftrightarrow (\text{ABp} \;\&\; \text{PR}(\text{p} \leftrightarrow \text{BAp}))$$

$$\vdash \text{ARp} \rightarrow \text{ABp}$$

$$\vdash \text{ARp} \rightarrow \text{COp}$$

$$\vdash \text{ARp} \rightarrow \text{PR}(\text{p} \leftrightarrow \text{BAp})$$

$$\vdash \text{ARp} \rightarrow (\text{p} \leftrightarrow \text{BAp})$$

$$\vdash \text{ARBAp} \rightarrow \text{ABBAp}$$

$$\vdash \text{AR}\neg\text{BAp} \rightarrow \text{CORFp}$$

$$\vdash \text{AR}\neg\text{BAp} \rightarrow \text{PR}(\neg\text{BAp} \leftrightarrow \text{RFp})$$

$$\vdash \text{AR}\neg\text{BAp} \rightarrow (\neg\text{BAp} \leftrightarrow \text{RFp})$$

It will be convenient to introduce an *agential reflectivity function* that associates with each world the set of propositions that are agentially reflective for our agent at that world:

$$AR_{\text{i}} = \{\text{X: X} \in CO_{\text{i}} \;\&\; \forall \text{j}[CO\text{ij} \rightarrow \text{j} \in BA\text{X} \cup (\text{W-X} \cap \text{W-}BA\text{X})]\}$$

So the set of agentially reflective propositions at i are those consistent with our agent's i-based abilities such that all worlds consistent with those abilities are either worlds where our agent brings about X (and hence X is true) or worlds where our agent does not bring about X and X is false.

It will also be convenient to introduce an operator for the notion of *a proposition's being open to an agent*, or just *propositional openness*:

$$\text{POp} =_{\text{df}} \text{ABp} \;\&\; \text{AB}\neg\text{p}$$

So that *it is Propositionally Open to an agent that p* if and only if the truth status of p is within the agent's current power. Given ABRE, this implies the indifference of propositional openness to negation:

$$\vdash \text{POp} \leftrightarrow \text{PO}\neg\text{p} \ \text{(IPO)}$$

It might be (and has been) thought that it is not really in my ability to bring a proposition about unless it is not predetermined that that proposition will be true, but this seems to be a mistake, for their can be redundancy of causal potential. It can surely be the case that *the only possible way* that I can stop someone else from doing something is if I do it myself first, as in our second phone call case. In such a case, the thing in question will occur no matter what, even though I could bring it about *myself* or not. This suggest another notion of openness, *agential openness*:

$$\text{AOp} =_{\text{df}} \text{POBAp}$$

So *it is Agentially Open to our agent that p* if and only if it is propositionally open to our agent that she brings it about that p. Here the notion is not p's status being open to an agent; rather it is the agent's agency with respect to p's truth that is open, as this trivially consequence indicates:

$$\vdash \text{AOp} \leftrightarrow (\text{ABBAp} \ \& \ \text{AB}\neg\text{BAp}).$$

AOp is perfectly compatible with PRp, but POp is not.

With this account of agency, ability, and inevitability in the background, we begin to develop a framework for aretaic notions.

Part II: The Aretaic Framework

Neutrality

Roughly, we can think of a neutral proposition as a proposition consistent with an agent's abilities but reflecting no positive or negative merit all-in-all on our agent. This means that our agent would not be worthy of praise nor blame for (the truth of) such a proposition. This can be true vacuously, because although the proposition is consistent with our agent's abilities, it is neither an available performance nor an agential reflection thereof, or non-vacuously, because although it is an available performance or reflection thereof, either it has no positive or negative aretaic value, or it has a perfect balance of positive and negative aretaic value, resulting in a "neutralization" of the two opposing values. To represent neutrality we introduce a *neutrality function*, $\mathcal{NL}$, that for each world, yields a set of propositions that are aretaically neutral for our imagined agent:

$$\mathcal{NL}\text{: W} \rightarrow \text{Pow(Pow(W)), that is, } \mathcal{NL}_i \subseteq \text{Pow(W)).}$$

The truth-conditions for the corresponding syntactic operator are:

$$\mathcal{M} \models_i \text{NLp iff } \|\text{p}\|^{\mathcal{M}} \in \mathcal{NL}_i.$$

Since we are interested especially in the positive and negative aretaic appraisal of performances consistent with our agent's ability, paving the way for linking such appraisal with the deontic appraisal of propositions within our agent's ability, we will require that neutral propositions be consistent with our agent's abilities:

$$\mathcal{NL}\text{-Confinement: } \forall i\, \forall X(X \in \mathcal{NL}_i \rightarrow X \in \mathcal{CO}_i)$$

This validates:

$$\text{NL-CO: } \models \text{NLp} \rightarrow \text{COp},$$

allowing us to derive:

$$\text{NL-OD: } \vdash \neg \text{NL}f$$

Since we will be centering our account of praiseworthiness and blameworthiness around neutrality, we need to consider what constraints to stipulate on neutrality, praiseworthiness and blameworthiness regarding their connections to agency. It is *agents* that are fundamentally worthy of praise and blame, and more specifically, it is fundamentally *for their agency* with respect to a given proposition that they are worthy of praise or blame. So, for non-neutral aretaic appraisal, we will restrict ourselves to performances and reflections thereof. Furthermore, we would like to evaluate not only the *actual* performances of agents, but ideally, past, present and future performances that are consistent with their abilities. Since in this preliminary work, we incorporate no resources for representing temporal matters, we are constrained to focusing on performances and reflections thereof that are consistent with our agent's *current* abilities. Let us introduce a function that identifies the set of *available performances*—set of performances that are consistent with our agent's abilities:

$$\mathcal{AP}_i = \{\mathcal{BA}X: X \in \mathcal{AB}_i\}$$

So for each world i, $\mathcal{AP}_i$ is the set of all propositions asserting of some X within our agent's ability, that our agent brings X about.

Truth conditions for the corresponding operator are:

$$\mathcal{M} \models_i \text{APp iff } \|p\|^{\mathcal{M}} \in \mathcal{AP}_i \text{ (i.e. } \exists X(\mathcal{BA}X = \|p\|^{\mathcal{M}} \ \& \ X \in \mathcal{AB}_i)$$

The following are valid:

$$\models \text{APp} \rightarrow \text{COp}$$

$$\models \text{ABp} \rightarrow \text{APBAp}$$

$$\models \text{ABp} \leftrightarrow \text{APBAp}$$

$$\models \text{COBAp} \leftrightarrow \text{APBAp},$$

$$\models \neg \text{AP}f$$

$$\text{If } \models p \leftrightarrow q \text{ then } \models \text{APp} \leftrightarrow \text{APq}$$

The fourth entry should not confuse us into thinking that the AP operator is equivalent to the CO operator. For the following are *both* valid:

$$\text{N-CO: } \models \text{CO}t$$

$$\text{AP-OD: } \models \neg \text{AP}t$$

What $\text{COBAp} \leftrightarrow \text{APBAp}$ indicates is that since BAp already has the explicit form of a performance, it follows from APBAp that p is within our agent's abilities, and conversely, if p is within our agent's abilities, then there is a proposition asserting that our agent brings about what p expresses that is consistent with our agent's abilities. So when the sentential complement of AP is itself in explicit agential form, the compound reduces to the result of replacing AP with CO. But when the sentential complement is not in agential form, but rather just denotes such a performance, primitively, as it were, then no such reduction is available, since we don't know what the proposition is that p asserts that our agent brings about.

We now stipulate the aretaic neutrality of all propositions consistent with our agent's ability that are neither available performances (directly agential) nor agentially reflective propositions (indirectly agential). Call this *the neutrality of consistent non-agent-indexed propositions*:

$$\mathit{NNAIP}: \forall \text{i} \, \forall \text{X}[(\text{X} \in \mathit{CO}_{\text{i}} \ \& \ \text{X} \notin \mathit{AP}_{\text{i}} \cup \mathit{AR}_{\text{i}}) \rightarrow \text{X} \in \mathit{NL}_{\text{i}}]$$

With *NNAIP*, the following is validated:

$$\text{NNAIP: } \models \text{COp} \rightarrow (\text{NLp} \vee \text{APp} \vee \text{ARp}),$$

which allows us to derive:

$$\vdash \text{COp} \leftrightarrow (\text{NLp} \vee \text{APp} \vee \text{ARp})$$

The *neutrality of necessary propositions* is now also derivable:

$$\text{NL-NEC: } \models \text{NL}t$$

However, it is one thing to say that there is always *some* proposition (W) that is aretaically neutral for an agent, but it is quite another to say that one can always act so that *only* neutral things occur. For some actions open to me can be praiseworthy to do and blameworthy to refrain from doing. Suppose you are a soldier in a group that, on a fair rotating basis, routinely places a sentry at a pivotal protective position with the understand that under an unexpected attack, whoever is there must hold that position until the group signals that it is properly prepared. In such a case, it is easy to imagine that it is praiseworthy to stay, blameworthy to refrain, and predetermined once the attack begins that I do one or the other. If so, no matter what I do, I do something non-neutral.

It might be thought that we should endorse another constraint:

$$\forall i\, \forall X(\forall j(\mathit{CO}ij \rightarrow j \in X) \rightarrow X \in \mathcal{NL}_i).$$

This validates the thesis that any proposition that is predetermined is neutral. But we should not rule out, as a matter of logic, the possibility that it can be predetermined that I bring something about, and in particular, something non-neutral. For one thing, compatibilists of the traditional sort, endorse the doctrine that all human action is pre-determined and that we are responsible for some of our actions. So presumably we will then be blameworthy and praiseworthy for many of those for which we are responsible. For another, it is not even clear that incompatibilists should endorse the underlying spirit of the above constraint, since it may be that prior genuine choices might make inevitable certain future choices that I am nonetheless to be blamed for, as when I negligently place myself in a situation of irresistible temptation.

RE is valid for NL:

$$\text{NL-RE: If } \models p \leftrightarrow q \text{ then } \models NLp \leftrightarrow NLq$$

Note that neutrality is not indifferent to negation,

$$NLp \leftrightarrow NL\neg p,$$

and thus is not to be construed as an indifference notion. The proposition ($\neg p$) that *I do not bring it about that I now do some wonderful thing* might be consistent with my ability, propositionally open and nonetheless neutral (I can do the same wonderful thing later, for example, with no loss). But that I do bring about something wonderful right now (p) needn't be neutral.

As already suggested, neutrality is consistent with conjunctions of positive and negative aretaic features. For example, consider the proposition that *I was rude to Bob (p) while I was helping Bill (q)*. Ordinarily, my rudeness to Bob would be blameworthy, and my helping Bill would be praiseworthy. But if the "amount" of positive and negative aretaic value is equal, then the conjunction will be neutral–neither praiseworthy nor blameworthy *all things considered*, even though there is clearly a praiseworthy thing and also a blameworthy thing implied by the conjunction. So we reject:

$$NL(p \ \& \ q) \rightarrow (NLp \ \& \ NLq).$$

Similarly, the fact that two propositions are individually neutral does not imply that their conjunction is. We reject:

$$(NLp \ \& \ NLq) \rightarrow NL(p \ \& \ q)$$

For one thing, there are propositions that can be neutral and have neutral negations, so this principle would imply that contradictions can be neutral. But even in cases of propositions jointly within my ability, it might be the case that although each is neutral, their conjunction will be blameworthy or praiseworthy. For example, suppose that I am culpably obligated to bring it about that either $\neg r$ or $\neg s$. Then where p is BAr and q is BAs, it might be neutral that p and neutral that q, yet not be neutral at all that p & q.

It might be thought that an RM principle for NL should hold:

$$\text{If} \vdash p \rightarrow q \text{ then } \vdash NLp \rightarrow NLq$$

That is, it might be thought that if p entails q and p is neutral then so is q, else the non-neutrality of q would infest p. There is a sense in which it is true that if q is non-neutral and p entails q, then the non-neutrality of q does impinge on p, but since we are dealing with all-in-all notions, it does not follow that p is thereby not all-in-all neutral. For example, consider the complex proposition *that I do something praiseworthy that aretaically counterbalances something I do that is blameworthy* (p). This implies *that I do something blameworthy* (q), but the proposition that implies it is paradigmatically neutral all-in-all.

Aretaic Preference and the Objects of Aretaic Appraisal

We first define a world-relative ordering function, which will yield a weak or quasi-ordering relation, $\geqslant_i$, for each world, i:

$$\geqslant: W \rightarrow Pow(Pow(W) \times Pow(W)), \text{ i.e. } \geqslant_i \subseteq Pow(W) \times Pow(W).$$

So for each world i, and pair of propositions, X and Y, $X \geqslant_i Y$ if and only if *X reflects as well on our agent as Y (X is aretaically as good as Y) from the standpoint of i.* A corresponding operator is introduced with the following truth-condition:

$$M \models_i p \geqslant q\text{: } \|p\|^M \geqslant_i \|q\|^M$$

It is clear that we sometimes take someone to be blameworthy or praiseworthy for *the results* of their performances. For example, we do say that a rescuer is to be praised for the fact that the threatened person was rescued. But since we wish to evaluate not only an agent's actual performances, but those consistent with her ability, we cannot say automatically that our agent would be praiseworthy for the threatened person's being saved. The problem here is that it may be the case that someone else brings about that proposition. If so, I can't be praiseworthy for the mere fact that it is true. So it can't be non-agential propositions within my ability that are the primary objects of positive (or negative) aretaic appraisal. It must be the performances consistent with my abilities that we evaluate primarily. Accordingly, we cannot say that your rescue should even be intuitively non-neutral for me, and ranked above the neutral propositions. My bringing your rescue about presumably would be praiseworthy for me, and someone else's rescuing you would be aretaically neutral for me, so the proposition that you are rescued per se could be neutral for me. However, if I am the only person in a position to make the rescue, that is, if the person's rescue reflects my agency, then we can say that I would be praiseworthy for the person's rescue. So we will concern ourselves with the aretaic evaluation of available performances and agential reflections thereof.

With this in mind, our $\geqslant_i$ rankings are intended primarily to rank performances consistent with our agent's current ability or reflections thereof, and only secondarily to rank, exclusively as neutral, propositions that are consistent with our agent's ability, but are non-agent-indexed. So we begin by constraining the relata of our ordering relations to propositions consistent with our agent's abilities:

$$\geqslant\text{-}CO \text{ Confinement: } \forall i(\geqslant_i \subseteq CO_i \times CO_i)$$

This rules out contradictions from entering into a ranking at any world, since no contradiction is consistent with an agent's ability:

We will also assume that all propositions consistent with our agent's abilities are self-comparable, and we will assume transitivity as well:

$$\text{Reflexive: } \forall i\,\forall X(X \in CO_i \rightarrow X \geqslant_i X)$$

$$\text{Transitive: } \forall i\,\forall X\,\forall Y\,\forall Z[(X \geqslant_i Y \;\&\; Y \geqslant_i Z) \rightarrow X \geqslant_i Z]$$

Clearly, confinement and reflexivity imply $\forall i\,\forall X(X \in CO_i \leftrightarrow X \geqslant_i X)$. We might thus designate the *aretaically evaluable* propositions, indifferently as those comparable with some proposition or other, those that are self-comparable or those consistent with our agent's abilities. However, we do *not* endorse

$$\geqslant\text{-Connectivity: } \forall i\,\forall X\,\forall Y[X, Y \in CO_i \rightarrow (X \geqslant_i Y \vee Y \geqslant_i X)],$$

as a basic constraint. In particular, it seems less clear that any two available performances must be aretaically comparable to one another. So I leave connectivity aside here, only briefly considering some consequences of adding it later on where it ties in with other matters.

A *strong preference* relation and an *equi-ranking* relation are definable:

$$X >_i Y =_{df} X \geqslant_i Y \;\&\; \neg(Y \geqslant_i X)$$

$$X =_i Y =_{df} X \geqslant_i Y \;\&\; Y \geqslant_i X$$

Similarly for the corresponding operators:

$$p > q =_{df} p \geqslant q \;\&\; \neg(q \geqslant p)$$

$$p = q =_{df} p \geqslant q \;\&\; q \geqslant p$$

Derivative truth-conditions for these operators are:

$$M \models_i p > q\text{: } \|p\|^M >_i \|q\|^M$$

$$M \models_i p = q\text{: } \|p\|^M =_i \|q\|^M$$

An agential reflection of an available performance and the performance it reflects ought to reflect equally on our agent. So we also stipulate the *reflective-ranking of agential reflections*:

$$AR = BA\text{: } \forall X(X \in AR_i, \text{ then } BAX =_i X)$$

The following basic schemata are validated:

$$\text{AR}=\text{BA}: \vDash \text{ARp} \rightarrow (\text{p}=\text{BAp})$$

$$\geqslant\text{-CO Confinement}: \vDash \text{p}\geqslant\text{q} \rightarrow (\text{COp} \ \& \ \text{COq})$$

$$\text{Rflx}(\geqslant): \vDash \text{COp} \rightarrow \text{p}\geqslant\text{p}$$

$$\text{Trans}(\geqslant): \vDash \text{p}\geqslant\text{q} \ \& \ \text{q}\geqslant\text{r}) \rightarrow \text{p}\geqslant\text{r}$$

The following basic rules will be validity-preserving:

$$\geqslant\text{-RE1: If } \vDash \text{p} \leftrightarrow \text{q then } \vDash \text{r}\geqslant\text{p} \rightarrow \text{r}\geqslant\text{q}$$

$$\geqslant\text{-RE2: If } \vDash \text{p} \leftrightarrow \text{q then } \vDash \text{p}\geqslant\text{r} \rightarrow \text{q}\geqslant\text{r}$$

From these axioms, rules, and our definitions, the following are derivable:

$$\geqslant\text{-}f: \vdash \neg(f\geqslant f)$$

$$\vdash \neg(f\geqslant \text{p} \vee \text{p}\geqslant f)$$

$$\text{Rflx}(=): \vdash \text{COp} \rightarrow \text{p}=\text{p}$$

$$\text{Trans}(=): \vdash \text{p}=\text{q} \ \& \ \text{q}=\text{r}\cdot\!\rightarrow \text{p}=\text{r}$$

$$\text{Sym}(=): \vdash \text{p}=\text{q} \rightarrow \text{q}=\text{p}$$

$$=\text{-CO Confinement}: \vdash \text{p}=\text{q} \rightarrow \text{COp} \ \& \ \text{COq}$$

$$>\text{-CO Confinement}: \vdash \text{p}>\text{q} \rightarrow \text{COp} \ \& \ \text{COq}$$

$$\text{Asym}(>): \vdash \text{p}>\text{q} \rightarrow \neg(\text{q}>\text{p})$$

$$\text{Tran}(>): \vdash \text{p}>\text{q} \ \& \ \text{q}>\text{r}\cdot\!\rightarrow \text{p}>\text{r}$$

$$\text{Mixed 1}: \vdash \text{p}\geqslant\text{q} \ \& \ \text{q}=\text{r}\cdot\!\rightarrow \text{p}\geqslant\text{r}$$

$$\text{Mixed 2}: \vdash \text{p}=\text{q} \ \& \ \text{q}\geqslant\text{r}\cdot\!\rightarrow \text{p}\geqslant\text{r}$$

$$\text{Mixed 3}: \vdash \text{p}>\text{q} \ \& \ \text{q}=\text{r}.\!\rightarrow \text{p}>\text{r}$$

$$\text{Mixed 4}: \vdash \text{p}=\text{q} \ \& \ \text{q}>\text{r}.\!\rightarrow \text{p}>\text{r}$$

$$>\text{-RE1 If } \vdash \text{p} \leftrightarrow \text{q then } \vdash \text{r}>\text{p} \rightarrow \text{r}>\text{q}$$

$$>\text{-RE2 If } \vdash \text{p} \leftrightarrow \text{q then } \vdash \text{p}>\text{r} \rightarrow \text{q}>\text{r}$$

$$=\text{-RE: If } \vdash \text{p} \leftrightarrow \text{q then } \vdash \text{p}=\text{r} \rightarrow \text{q}=\text{r}$$

Given our picture, we must add a constraint, *the equality of all neutral propositions*:

$$\mathit{ENL}: \forall \text{i} \, \forall \text{X} \, \forall \text{Y}[\text{X} \in \mathcal{NL}_{\text{i}} \rightarrow (\text{Y} \in \mathcal{NL}_{\text{i}} \leftrightarrow \text{X} =_{\text{i}} \text{Y})]$$

This validates:

$$\text{ENL}: \vDash \text{NLp} \rightarrow (\text{NLq} \leftrightarrow \text{p}=\text{q})$$

The *equality of neutral with necessary propositions* is derivable:

$$\text{ENPNP: } \vdash \text{NLp} \rightarrow \text{p} = t.$$

Similarly for

$$\begin{aligned} \vdash \text{NLp \& NLq} \cdot\!\rightarrow\!\cdot \; & \text{a) } (\text{r} \geqslant \text{p}) \rightarrow (\text{r} \geqslant \text{q}) \ \& \\ & \text{b) } (\text{p} \geqslant \text{r}) \rightarrow (\text{q} \geqslant \text{r}) \ \& \\ & \text{c) } (\text{r} > \text{p}) \rightarrow (\text{r} > \text{q}) \ \& \\ & \text{d) } (\text{p} > \text{r}) \rightarrow (\text{q} > \text{r}), \end{aligned}$$

and,

$$\begin{aligned} & \vdash \text{p} > \text{q} \rightarrow (\neg \text{NLp} \vee \neg \text{NLq}) \\ & \vdash \text{p} > \text{q} \rightarrow (\text{APp} \vee \text{APq} \vee \text{ARp} \vee \text{ARq}) \\ & \vdash \text{p} > t \rightarrow (\text{APp} \vee \text{ARp}) \\ & \vdash t > \text{p} \rightarrow (\text{APp} \vee \text{ARp}) \\ & \vdash (\text{COp} \ \& \ \neg \text{NLp}) \rightarrow (\text{APp} \vee \text{ARp}) \\ & \vdash \text{AR} \neg \text{BAp} \rightarrow (\text{NL} \neg \text{BAp} \leftrightarrow \text{NLRFp}) \end{aligned}$$

Note that the neutrality of non-agent-indexed propositions introduced earlier, along with $\geqslant$-*CO* Confinement, the equality of neutral propositions and the neutrality of necessary propositions implies that positive and negative aretaic appraisal will be confined to agential performances and reflections thereof:

$$\forall \text{i} \, \forall \text{X}[(\text{X} >_{\text{i}} \text{W} \vee \text{W} >_{\text{i}} \text{X}) \rightarrow (X \in AR_{\text{i}} \vee \text{X} \in AP_{\text{i}}).$$

So the positive or negative ranking of propositions is restricted to performances consistent with our agent's ability or to reflections thereof. This is plausible, since such a proposition is intended to reflect positive or negative merit on an agent for her *performance*. Since we are not using temporal resources, in our freeze-frame setting, we must imagine that aretaic evaluation is also confined to what is consistent with our agent's *current* ability. So we must imagine that "the aretaic slate has been cleaned", or that we are here concerned with praise, blame and neutrality, based only on our agent's current abilities. With the addition of temporal resources, a more sensitive treatment would no doubt be possible.

We will consider some further secondary constraints below, in the context of discussing the concepts of positive and negative aretaic appraisal.

Indifference

Some propositions will be *aretaically indifferent* for our imagined agent. We define this notion as follows:

$$\text{AIp} =_{\text{df}} \text{NLp} \ \& \ \text{NL}\neg\text{p}$$

So *it is Aretaically Indifferent for our agent that p* if and only if both p and ¬p are neutral for our agent. As stated earlier indifference should be stronger than mere neutrality.

The derivative truth-condition for AI is:

$$M \models_{\text{i}} \text{AIp iff } \|\text{p}\|^{M} \in \mathcal{NL}_{\text{i}} \ \& \ \text{W-}\|\text{p}\|^{M} \in \mathcal{NL}\text{i}.$$

The indifference of indifference to negation, the equality of indifferent propositions, the equality of indifferent propositions to their negations, the non-indifference of necessary and impossible propositions, the equality of indifferent propositions with neutral propositions, the consistency of indifference with our agents' abilities, and an RE rule for AI are derivable:

$$\text{IAIN: } \vdash \text{AI} \leftrightarrow \text{AI}\neg\text{p}.$$

$$\text{EAI: } \vdash (\text{AIp} \ \& \ \text{AIq}) \rightarrow \text{p} = \text{q}.$$

$$\text{EIPN: } \vdash \text{AIp} \rightarrow \text{p} = \neg\text{p}$$

$$\text{AI-NO: } \vdash \neg\text{AI}t$$

$$\text{AI-OD: } \vdash \neg\text{AI}f$$

$$\text{ENL-AI: } \vdash \text{AIp} \rightarrow (\text{NLq} \leftrightarrow \text{p} = \text{q})$$

$$\text{AI-CO: } \vdash \text{AIp} \rightarrow \text{COp}.$$

$$\text{AI-RE: If } \vdash \text{p} \leftrightarrow \text{q then } \vdash \text{AIp} \leftrightarrow \text{AIq}.$$

Praiseworthiness and Blameworthiness

As alluded to earlier, we take the praiseworthy (blameworthy) propositions as those ranked aretaically higher (lower) than some neutral propositions, and this idea is indirectly captured by these concise definitions:

$$\text{PWp} =_{\text{df}} \text{p} > t$$

$$\text{BWp} =_{\text{df}} f > \text{p}$$

The derivative truth conditions are:

$$M \models_{\text{i}} \text{PWp iff } \|\text{p}\|^{M} >_{\text{i}} \text{W}$$

$$M \models_{\text{i}} \text{BWp iff W} >_{\text{i}} \|\text{p}\|^{M}$$

The following principles are validated and derivable:

PW-BW EXCL: $\vdash PWp \rightarrow \neg BWp$

PW-NL EXCL: $\vdash PWp \rightarrow \neg NLp$

BW-NL EXCL: $\vdash BWp \rightarrow \neg NLp$

PW > BW: $\vdash (PWp \ \& \ BWq) \rightarrow p > q$

PW > NL > BW: $\vdash (PWp \ \& \ NLq \ \& \ BWr) \rightarrow (p > q \ \& \ q > r)$

$\vdash NLp \rightarrow (PWq \leftrightarrow q > p)$

$\vdash NLp \rightarrow (BWq \leftrightarrow p > q)$

$\vdash ARp \rightarrow (PWp \leftrightarrow PWBAp)$

$\vdash ARp \rightarrow (BWp \leftrightarrow BWBAp)$

$\vdash AR\neg BAp \rightarrow (BW\neg BAp \leftrightarrow BWRFp)$

$\vdash AR\neg BAp \rightarrow (PW\neg BAp \leftrightarrow PWRFp)$

The following theorems reflect constraints on items ranked above or below the neutral zone:

PW-CO: $\vdash PWp \rightarrow COp$

BW-CO: $\vdash BWp \rightarrow COp$

$\vdash PWp \rightarrow (APp \vee ARp)$

$\vdash BWp \rightarrow (APp \vee ARp)$

PWBA-AB: $\vdash PWBAp \rightarrow ABp$

BWBA-AB: $\vdash BWBAp \rightarrow ABp$

The following indifference exclusion principle is also derivable:

AI-EXCL:

$$\vdash AIp \rightarrow (COp \ \& \ \neg BWp \ \& \ \neg BW\neg p \ \& \ \neg PWp \ \& \ \neg PW\neg p)$$

The only thing that blocks the converse of AI-EXCL,

AI-EXCL′: $(COp \ \& \ \neg BWp \ \& \ \neg BW\neg p \ \& \ \neg PWp \ \& \ \neg PW\neg p) \rightarrow AIp$

is incomparability. There may be some performances consistent with our agent's ability, and thus aretaically evaluable, that are not comparable to *t*, and so not "placed" above, below, or among the neutrals:

$$COp \rightarrow (NLp \vee PWp \vee BWp)$$

Such propositions would presumably contain conflicting positive and negative aretaic components pulling above and below the neutral line in a way that doesn't allow resolution according to our ordering. If, however, we endorsed

$$\geqslant\text{-Connectivity: } \forall i\, \forall X\, \forall Y[(X, Y \in CO_i) \rightarrow (X \geqslant_i Y \vee Y \geqslant_i X)],$$

we validate comparability:

$$\text{CO-COMP: } \models (\text{COp} \ \& \ \text{COq}) \rightarrow (p \geqslant q \vee q \geqslant p)$$

Given connectivity and CO-COMP, the following are validated and derivable:

$$\text{CO-COMP}': \vdash (\text{COp} \ \& \ \text{COq}) \rightarrow (p > q \vee q > p \vee p = q)$$

$$\text{AB-COMP}: \vdash (\text{ABp} \ \& \ \text{ABq}) \rightarrow (\text{BAp} \geqslant \text{BAq} \vee \text{BAq} \geqslant \text{BAp})$$

$$\text{CO-DEF}'': \vdash \text{COp} \leftrightarrow (p \geqslant t \vee t \geqslant p)$$

$$\vdash \text{COp} \leftrightarrow (p > t \vee t > p \vee p = t)$$

$$\vdash \text{COp} \leftrightarrow (\text{NLp} \vee \text{PWp} \vee \text{BWp})$$

$$\text{AB-DEF}': \vdash \text{ABp} \leftrightarrow (\text{BAp} \geqslant t \vee t \geqslant \text{BAp})$$

$$\vdash \text{ABp} \leftrightarrow (\text{BAp} > t \vee t > \text{BAp} \vee \text{BAp} = t)$$

$$\text{NL-DEF}': \vdash \text{NLp} \leftrightarrow (\text{COp} \ \& \ \neg \text{BWp} \ \& \ \neg \text{PWp})$$

$$\text{AI-EXCL}': \vdash (\text{COp} \ \& \ \neg \text{BWp} \ \& \ \neg \text{BW} \neg p \ \& \ \neg \text{PWp} \ \& \ \neg \text{PW} \neg p) \rightarrow \text{AIp}$$

$$\text{AI-DEF}': \vdash \text{AIp} \leftrightarrow (\text{COp} \ \& \ \neg \text{BWp} \ \& \ \neg \text{BW} \neg p \ \& \ \neg \text{PWp} \ \& \ \neg \text{PW} \neg p)$$

$\geqslant$-Connectivity validates COMP, and as the above indicate, with COMP, we are able to define CO, NL, and AI via our ordering concepts. This suggests exploring the possibility of similar reductions without COMP.

We now turn to a few more subtle issues concerning praiseworthiness, blameworthiness, agency, and comparability.

One question we might ask about the concept of blameworthy performances is do they satisfy this no conflicts principle:

$$\text{BWBA-NC: BWBAp} \rightarrow \neg \text{BWBA} \neg p.$$

Suppose POp. Now, there are many things that I have no business bringing about whose negation I also have no business bringing about. For example, consider parental interference. Ordinarily, I have no business bringing it about that your child is reprimanded (p) and I also have no business bringing it about that your child is not reprimanded (i.e. by anyone). That's up to you, not me. I would be blameworthy if I took either option. But I can take a third option, to not interfere, since AOp.

Similar remarks apply to praiseworthy performances:

$$\text{PWBA-NC: PWBAp} \rightarrow \neg\text{PWBA}\neg\text{p}.$$

Suppose again that although POp, p is not currently in my jurisdiction. Now imagine that whoever places the status of p in his jurisdiction places himself at avoidable and unrequired risk. It is not hard to imagine that, in such a case, it might be praiseworthy for me to bring about p or to bring about $\neg$p.

Notice, however, that neither case refutes the following:

$$\text{BW-NC}': \text{BWBAp} \rightarrow \neg\text{BWRFp}$$

$$\text{PW-NC}': \text{PWBAp} \rightarrow \neg\text{PWRFp}.$$

So what of these? (I assume we do not read RF as logically implying that I am *tempted* to bring about p.) It might be thought that since one can have conflicts of obligation, it is obvious that one can have conflicts of blameworthiness. But it is not clear that this follows. Suppose that because I am negligent, I make two appointments for the same time, and realize afterward that it is too late to cancel either one. Add that there is no more reason to favor one particular appointment over the other. As is natural, assume that my not attending meeting 1 is agentially reflective. Similarly for my not attending meeting 2. Intuitively, I am obligated to go to the first appointment, and I am obligated to go to the second appointment. But since the dilemma is due to negligence, we can easily imagine that I am *culpably obligated* to do these things: I am obligated to go to appointment 1 (2), and I am blameworthy for refraining from going to appointment 1 (2). Let's represent this situation as follows, using OB for "it is obligatory that", and CB for "it is culpably obligatory that":

CBBA(I go to appointment 1),

CBBA(I go to appointment 2),

$\neg$AB(I go to appointment 1 & I go to appointment 2).

AR$\neg$BA(I go to appointment 1)

AR$\neg$BA(I go to appointment 2)

So OBBA(I go to appointment 1),

OBBA(I go to appointment 2),

BWRF(I go to appointment 1),

BWRF(I go to appointment 2),

BW$\neg$BA(I go to appointment 1),

BW$\neg$BA(I go to appointment 2).

Here we have a classic conflict of obligations (enriched in ways that suggest some of the interest in a deontic expansion of our scheme). The crucial question for us here is, would I be blameworthy for *keeping* appointment 1 because it results in my not going to appointment 2, which is blameworthy? I do not think this follows. I think I would not be blameworthy for *going to* either appointment, since after all, that is *precisely* what I have a primary obligation to do, and what I am culpable for *not doing*. So here, we do not seem to wind up with a case where it is blameworthy for me to bring it about that I am at appointment 1, even though bringing about my attendance at appointment 1 results in my missing 2, which is blameworthy. The blame just does not seem to clearly distribute the way the argument envisions.

However, suppose we were to change the case slightly so that it is more blameworthy to not go to appointment 2 than to not go to appointment 1 (say because 2 is plainly more important than 1), even though either is still blameworthy to some extent. Then it is at least plausible to say that it would be blameworthy to attend appointment 1 in such a case. (This suggests exploring the question of whether or not attending 1 and not attending 1 could be *equally blameworthy*. I suspect we can only get an intuitively valid rendering in an enriched setting, perhaps with temporal operators as well.)

For these reasons, I have not treated BW-NC′ or PW-NC′ as core axioms, but since they have an (at least) initially plausible ring, I identify the constraints that validate them:

$$BW\text{-NC}': \forall \mathrm{i}\, \forall \mathrm{X}(t >_{\mathrm{i}} BA\mathrm{X} \rightarrow \neg(t >_{\mathrm{i}} BA(\mathrm{W}\text{-}BA\mathrm{X}))).$$

$$PW\text{-NC}': \forall \mathrm{i}\, \forall \mathrm{X}(BA\mathrm{X} >_{\mathrm{i}} t \rightarrow \neg(BA(\mathrm{W}\text{-}BA\mathrm{X}) >_{\mathrm{i}} t)).$$

BW-NC tells us that for any world i, and performance *BA*X, if *t* is ranked higher than *BA*X, then *t* is not ranked higher than bringing about the negation of *BA*X. *PW*-NC gives us the mirror image. These two validate the following *upper and lower exclusion* principles:

$$t{>}\mathrm{EXCL}: \vDash t > \mathrm{BAp} \rightarrow \neg(t > \mathrm{RFp})$$

$$t{<}\mathrm{EXCL}: \vDash \mathrm{BAp} > t \rightarrow \neg(\mathrm{RFp} > t)$$

These are then derivable:

$$\text{BW-NC}': \quad \vdash \mathrm{BWBAp} \rightarrow \neg \mathrm{BWRFp}$$

$$\text{BW-NC}': \quad \vdash \mathrm{PWBAp} \rightarrow \neg \mathrm{PWRFp},$$

$$\vdash \mathrm{BAp} = \mathrm{RFp} \rightarrow \cdot \neg(t > \mathrm{BAp}) \;\&\; \neg(t > \mathrm{RFp})$$

$$\vdash \mathrm{BAp} = \mathrm{RFp} \rightarrow \cdot \neg(\mathrm{BAp} > t) \;\&\; \neg(\mathrm{RFp} > t)$$

$$\vdash \mathrm{BAp} = \mathrm{RFp} \rightarrow (\neg \mathrm{BWBAp} \;\&\; \neg \mathrm{BWRFp})$$

$$\vdash \mathrm{BAp} = \mathrm{RFp} \rightarrow (\neg \mathrm{PWBAp} \;\&\; \neg \mathrm{PWRFp}).$$

With COMP as well, we can derive:

$$\vdash \mathrm{BAp} = \mathrm{RFp} \rightarrow (\mathrm{NLBAp}\ \&\ \mathrm{NL}\neg\mathrm{BAp}\ \&\ \mathrm{NLRFp}).$$

In considering the moral dilemma above, I alluded to the concept of something's being more blameworthy than some other thing. Although it is beyond the scope of this study to pursue these, let me note these suggestive analyses of some salient comparative aretaic notions:

p is as praiseworthy as q: $p \geqslant q$ & PWp

p is more praiseworthy than q: $p > q$ & PWp

p is less praiseworthy than q: $q > p$ & PWq

p is just as praiseworthy as q: $p = q$ & PWp

p is as blameworthy as q: $q \geqslant p$ & BWp

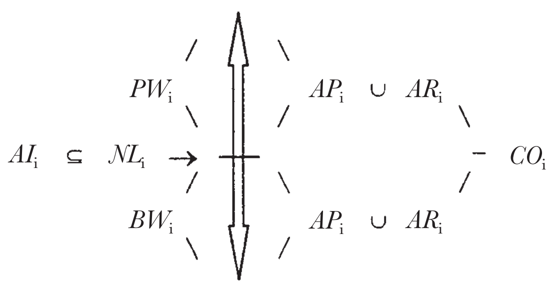
p is more blameworthy than q: $q > p$ & BWp

p is less blameworthy than q: $p > q$ & BWq

p is just as blameworthy as q: $p = q$ & BWp

The following diagram reflects some fundamental elements of our aretaic framework, and its dependency on the agential notions (it assumes $\geqslant$-connectivity for simplicity):

Location of: Confined to:



Conclusion

The framework for agency, ability and inevitability, with minimal models for BA and Kripke models for PR, warrants further exploration. in its own right, including exploration of iterated and embedded operators. The framework for the aretaic notions is adaptable to other families of evaluative concepts, in some cases more easily, since less restrictively.

Regarding our combined agential-aretaic framework, some interesting conceptual reductions are possible both via, and within, the aretaic component. For $\geqslant$-CO Confinement and $\geqslant$-Reflexivity imply $\forall i\,\forall X(X \in CO^{*}_{i} \leftrightarrow X \geqslant_{i} X)$. But then we might identify the propositions consistent with our agent's abilities with those that are *aretaically evaluable*, defining CO as: $COp =_{df} p \geqslant p$. Since all notions in the predetermination module are definable via CO, the notions in that module are all definable via our ordering notion. Furthermore, the *equality of neutral with necessary propositions* is also derivable, as is its converse. Thus neutrality is equivalent to equi-ranking with a tautology: $\vdash NLp =_{df} p = t$. So it appears that only BA, $\geqslant$, and AP need be taken as basic, and the only semantic notions required beyond set theory, are just *W*, *BA* and $\geqslant$. This is interesting in itself, and it may facilitate metatheory.

Finally, adding a deontic module for obligation should be interesting. As suggested by our example of a moral dilemma, it would allow us to distinguish *culpable and non-culpable violations* of obligations, and it would allow us to represent *various connections ethicists have endorsed* between what is obligatory, permissible and impermissible, and what is praiseworthy and blameworthy. Also, adding even a simple deontic module for obligation would allow for the formulation of most accounts by ethicists of supererogation and action beyond the call of duty (but not that in McNamara 1996). Adding it to McNamara's "Doing Well Enough" framework would allow us to distinguish supererogation from action beyond the call of duty, among other things. These additions would take us one step closer to modeling common sense morality, thereby adding some support for its coherence under scrutiny, and hence for its legitimacy.

Acknowledgments

I wish to thank David Makinson and Leon van der Torre for useful critical conversation about earlier preliminary Andersonian-Kangerian models for praiseworthiness and blameworthiness, and for gently prodding me to look for better structures than those. Thanks also go to Jose Carmo, Alexandra Deegan, Andrew Jones, and Luis Nagel for beneficial conversations about the general approach herein.

References

Brown, M. 1992. Normal bimodal logics of ability and action. *Studia Logica*, vol. 2, pp. 519–532.

Brown, M. and Carmo, J. 1996. *Deontic Logic, Agency and Normative Systems: DEON '96.* Springer-Verlag, Berlin.

Chellas, B. 1980. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge.
Chisholm, R. and Sosa, E. 1966. On the logic of 'intrinsically better'. *American Philosophical Quarterly*, vol. 3, pp. 244–249.
Elgesem, D. 1993. *Action Theory and Modal Logic*. Dissertation. Department of Philosophy, University of Oslo.
Elgesem, D. 1997. The modal logic of agency. *Nordic Journal of Philosophical Logic*, vol. 2, no. 2, pp. 1–46.
Holmstrom-Hintikka, G. and Tuomela, R. (eds.) 1997. *Contemporary Action Theory*, vol. II. Kluwer Academic Publishers, Dordrecht.
McNamara, P. 1996. Must I do what I ought (or will the least I can do do)? In Brown and Carmo, 1996, pp. 154–173.
McNamara, P. 1999. Toward an integrated agential, aretaic and deontic framework. Manuscript.
Santos, F. and Carmo, J. 1996. Indirect action, influence and responsibility. In Brown and Carmo, 1996, pp. 194–215.
Santos, F., Jones, A. and Carmo, J. 1997. Responsibility for action in organizations: a formal model. In Holmstrom-Hintikka and Tuomela, 1997, pp. 333–350.

Department of Philosophy,
University of New Hampshire,
Durham, NH, USA
E-mail: paulm@cisunix.unh.edu