**Norms**

Norms are standards of behaviour that are not obviously subjectively prudential in nature. Typically, they are presented as rules that put limits on the kinds of actions that can result from the practical reasoning of the agent. For example, actions which fall outside the range permitted by the appropriate norms are not taken *even where it is estimated that, were no such norms to exist, taking them would lead to greater total satisfactions than any other action*.

Although the fact that gives these rules the property of being norms is disputed (as is the more philosophical notion of normativity in general – which does not concern us here) it does seem to be the case that norms are quite simply related to the actual circumstances of their operations and the actions that they forbid, permit, or mandate. The relationship in the latter case is straightforwardly logical: the forbidden action is logically inconsistent with the norm, the permitted action is consistent with it, and the mandated action is the *only* action consistent with it. Taking this cue we develop the following definitions.

***Mode of Action***

A Norm is an action-directing rule, *n*, which may be positive or negative.

1. For a Positive Norm:

(*Cx,applies*  *x*) (*x,good* *x*) (*cx,applies*  *Cx,applies*) (*ax*  *x,good*)

[*cx* = *cx,applies*  *n* & *cx* & *ax* ⊢  ]

» *There is a class of relevant contexts and of acceptable actions for an agent therein such that if an agent is in any of those relevant contexts then if the action is not one of the acceptable ones then the action in that context is logically inconsistent with the norm.*

1. For a Negative Norm:

(*Cx,applies*  *x*) (*x,bad* *x*) (*cx,applies*  *Cx,applies*) (*ax*  *x,bad*)

[*cx* = *cx,applies*  *n* & *cx* & *ax* ⊢  ]

» *There is a class of relevant contexts and unacceptable actions for an agent such that if an agent is in any of those relevant contexts then if the action is one of the unacceptable ones then the action in that context is logically inconsistent with the norm.*

* By the standard understanding of an intentional agent, logical reasoning is at least a part of the process by which *x* goes about determining his action in the circumstances, though we cannot depend upon it being a main part or that it will be very accurate.
* Note that the norm itself defines the relevant classes of contexts and actions.

We say that *cx* is a Context Governed By *n* for the norm *n* and write *g*(*n*, *cx*) iff

*cx*  *Cx,applies*

» *cx is a context for which the norm, n, is relevant.*

* i.e. if n is a negative norm:

(*Cx,applies*  *x*) (*x,bad* *x*)

[ (*cx,applies*  *Cx,applies*)(*ax*  *x,bad*)[*cx* = *cx,applies*  *n* & *cx* & *ax* ⊢ ] &

*cx*  *Cx,applies* ]

and if *n* is a positive norm

(*Cx,applies*  *x*) (*x,good* *x*)

[ (*cx,applies*  *Cx,applies*)(*ax*  *x,good*)[*cx* = *cx,applies*  *n* & *cx* & *ax* ⊢  ] &

*cx*  *Cx,applies* ]

* The distinction between positive and negative norms can be ignored in most cases

We say that *ax* is an Action Governed In *cx* By *n* for the context *cx* and the norm *n*

and write *g*(*n*, *cx*, *ax*) iff

*cx*  *Cx,applies* & *ax* = *Ax*(*qx*, *cx*)

» *The context is governed by n and the action is produced in that context.*

* i.e. if *n* is a negative norm :

(*Cx,applies*  *x*) (*x,bad* *x*)

[ (*cx,applies*  *Cx,applies*) (*ax*  *x,bad*) [*cx* = *cx,applies*  *n* & *cx* & *ax* ⊢ ] &

(*cx*  *Cx,applies* & *ax* = *Ax*(*qx*, *cx*)) ]

and if *n* is a positive norm

(*Cx,applies*  *x*) (*x,good* *x*)

[ (*cx,applies*  *Cx,applies*)(*ax*  *x,good*)[*cx* = *cx,applies*  *n* & *cx* & *ax* ⊢  ] &

(*cx*  *Cx,applies* & *ax* = *Ax*(*qx*, *cx*)) ]

* Again, the distinction between positive and negative norms can be ignored in most cases
* Note that several norms may govern any action in a context
* Let *aX* = {*ax1*, *ax2*, …, *axn*} be actions of agents *x1*, *x2*, …, *xn* in *X* in the context *cX*; then

the norm *n* governs *aX* in *cX*, *g*(*n*, *cX*, *aX*) iff (*axi*  *aX*) [*g*(*n*, *cX*, *axi*)]

Let *n* be a norm, *cx* a context with *g*(*n*, *cx*).

We say that action *ax*  *x* is Permitted by *n* in *cx* and write *n,cx*(*ax*) iff:

~(*n* & *cx* & *ax* ⊢ 

» *If a context is governed by n then the action in that context is logically consistent with the norm.*

We say that *ax*  *x* is Mandated by *n* in *cx* and write *n,cx*(*ax*) iff:

*~**n,cx*(*~ax*)

 *n* & *cx* & ~*ax* ⊢ 

 (*ax’*  *ax*) [*n* & *cx* & *ax’* ⊢ 

» *If a context is governed by n then any action other than the action ax in that context is logically inconsistent with the norm.*

* It probably isn’t necessary to distinguish the positive and negative norms in the modal notation, but it could be done by using +*n,cx* and -*n,cx*, +*n,cx* and -*n,cx*.

A Norm Formation is a set of norms, *N* = {*n1*, …, *nn*}.

* Standard classes of norm formations are conventions, traditions, etc. (which will be analyzed later.)
* Let *A* = {*ax1*, *ax2*, …, *axn*} be a set of actions by agents *x1*, *x2*, …, *xn*, in the corresponding contexts *C* = {*cx,1*, *cx2*, …, *cxn*}

The norm formation *N* governs *A* in those contexts, *g*(*N*, *C*, *A*), iff

(*axi*  *A*) (*ni*  *N*) [*g*(*ni*, *cxi*, *axi*)]

|  |
| --- |
| **Example:**An example of a simple (negative) norm might be:(*cx* *Cx,war* *Cx,self-defense*) (*ax*  *x,kill*) [~*ax*]which would be the form of the general statement that ‘Except in war or in self-defence, thou shalt not kill.’Consider the case that might apply if *x* has been insulted. The reasoning could include1. *cx* *Cx,insult*
2. (*ax*  *x,kill*)(*ax’* *ax*) [*T*(*C*(*ax*, *cx*), *Ix*) > *T*(*C*(*ax’*, *cx*), *Ix*)]
3. *ax* (by i. & ii.)

which is to say that the decision would be made to kill. But if the reasoning included the norm as sketched above, we would have 1. *cx* *Cx,insult*
2. (*ax*  *x,kill*)(*ax’* *ax*) [*T*(*C*(*ax*, *cx*), *Ix*) > *T*(*C*(*ax’*, *cx*), *Ix*)]
3. *ax* (by i. & ii.)
4. (*cx* *Cx,war* *Cx,self-defense*) (*ax*  *x,kill*) [~*ax*]
5. *Cx,insult*  (*Cx,war* *Cx,self-defense*) = 
6. *cx*  *Cx,war* *Cx,self-defense*
7. ~*ax* (by iv. & vi.)
8.  (by iii. & vii.)

Meaning that the decision to kill is impermissible |

Note that many definitions or descriptions of norms either explicitly or implicitly make them a characteristic of a ‘group’ (however defined,) or of a ‘subgroup,’ or of a society. A typical example is the claim (by Schaefer, R. T., *Sociology: A Brief Introduction*, 9th ed. McGraw-Hill, 2010, p. 65) that norms are “the established standards of behaviour maintained by a society.” Certainly, norms are only *sociologically* interesting when they are held by a collection of agents, and give rise to expectations of behaviour amongst the members of that collection; however, consistent with our ACS approach, we accept that they can only have a social effect by being implemented at the agent level. The definition given here makes no explicit reference to any social context, nor would it seem to be necessary to modify it to do so in order to bring it into conformity with that usage. At worst, the current definition would be of a superset of the sociologically interesting norms, but nothing significant (regarding generation, transmission, acceptance, etc. of norms) seems to hang on the distinction that would make social context *constitutive* of a norm. We may accept, therefore, that the term as defined both minimizes vagueness and satisfactorily interprets the term as commonly used.

The proper explanation for an agent’s adherence to norms is disputed. Bicchieri & Muldoon (2011, *sv*. ‘Social Norms’ in Zalta, (ed.) *SEP* (2014), §3) indicate that there are three principal theoretical approaches to the problem: the socialized actor, according to which norms are observed because they have been internalised to form part of the motivating character of the agent; the social identity theory, which explains it in terms of a preference for observing norms associated with an identity; and the RCT, which appeals to the expected total satisfaction to be gained given the expectation that norms will be enforced. Note that these are classes of theories, or perspectives on a problem, rather than well-defined particular theories. Note also that there is no good reason – independent of the claims of some HLST – to take these as mutually exclusive. In fact there are reasonably straightforward ways to incorporate each of those perspectives into the ACS framework. We shall consider at this time only the cases of the socialized actor and the RCT (as it is generally understood.) For the social identity perspective, see later in the section on ‘Identities.’

***Internalized Norms***

The *locus classicus* for the socialized actor theory of norms is the work of T. Parsons. According to Parsons (1951, *The Social System*, New York: Routledge) “people voluntarily adhere to the shared value system because it is introjected to form a constitutive element of the personality itself.” In the vocabulary we have developed so far this is the claim that adherence in action to norms (*qua* norms) is a consequence of the *quality* of the agent. There are doubtless many ways in which this can be treated. We consider just two here.

1. *Norms as Beliefs*

One possibility is that they correspond to a certain class of beliefs that are important in the subjective determination of interest-maximising behaviour in that they impose restrictions on the kinds of actions that the agent believes are permissible, forbidden, or compulsory in certain circumstances. Such beliefs are beliefs about norms of action. In defence of this interpretation, it can be said that any agent bound by such a norm could be said to have an effective or implicit belief to that effect, and a natural understanding of an agent’s relationship to norms treats them as propositions featuring in the inferential processes contributing to a process of practical reasoning. It is possible to model that understanding as follows.

An agent *x* B-Accepts the norm *n* iff *x* believes the rule *n* to be true; i.e. iff *Bx*[ *n* ].

* Note that the assumption here is that norms are in propositional form. This seems reasonable (see example above,) but could be stipulated wolog.
* In this case, there is a condition for accepting that *Bx*[ *n* ] based on the notion that we can discover whether *x* *really* accepts a norm by observation of his behaviour. Thus for a negative norm:

*Bx*[ *n* ]  (*Cx,applies*  *x*) (*x,bad* *x*) (*cx,applies*  *Cx,applies*) (*ax*  *x,bad*) (*cx*  *x*)

[(*cx* = *cx,applies*  *n* & *ax* ⊢ )  *Ax*(*qx*, *cx*)  *ax*]

» *If someone produces an action in a circumstance in which the norm determines that that action is impermissible, then they cannot be said to accept that norm. The same applies* m.m. *for positive norms.*

* The *reasons* why *x* might fail this test are many. It might just be that *x* is not very clever and fails to reason accurately; perhaps the norm is too complex for any human; perhaps information is lacking; etc. We do not need to assume that *x* is actually dishonest or deluded, though those are also possibilities.
* It would seem, on the face of it anyway, that it’s quite possible that ~*Bx*[ *n* ] & *Bx*[ *Bx*[ *n* ]]; and other curiosities of epistemic logic may also apply. It will be necessary eventually to decide which axiomatization of epistemic logic is most appropriate for the operator *Bx*. We can ignore the question for now.

For later use we introduce the following terms.

An agent *x* is an B-Acceptant of the norm *n* iff *x* B-accepts the rule *n*.

The B-Acceptate of the norm *n*, *KB*(*n*), is all the agents that B-accept *n*. We will write

*KB*(*n*) = {*x*: *Bx*[ *n* ]}

* An agent *x* is a B-acceptant of the norm formation *N* iff *x* B-accepts the rules in *N*.
* The B-acceptate of the norm formation *N* is the set of all agents who B-accept the rules in *N*.

*KB*(*N*) = {*x*: *Bx*[ *N* ]}

*Norm Fuzziness*

The condition given for *x* being a B-acceptant of *n* may not adequately reflect the fuzziness of our belief states, or the degrees to which we believe something. Assuming that we wish to do so, and that this cannot be done by massaging the context variable of the action function as it applies to the conditions under which we will apply the norm, we may propose the following refinement for negative norms (and, again, *m*.*m*. for positive norms:)

*Bx*[ *n* ]  (*Cx,applies*  *x*) (*x,bad* *x*) (*cx,applies*  *Cx,applies*) (*ax*  *x,bad*) (*cx*  *x*)

[((*cx* = *cx,applies*  *n* & *ax* ⊢ ) &

~*Bx­*[ (*ax,alt* *ax*)[*T*(*C*(*ax,alt*, *cx*), *Ix*) << *T*(*C*(*ax*, *cx,*), *Ix*)] ])

 *Ax*(*qx*, *cx*)  *ax*]

» *No one will contravene n unless he estimates that the total satisfaction to be gained from doing so is sufficiently greater than the satisfaction to be gained from abiding by n.*

* It is intended that this should be able to account also for the distinction made between weakly and strongly enforced norms, and (perhaps equivalently) important and unimportant norms. An example of the latter distinction would be the norm of not murdering versus the norm of not wiping one’s nose on one’s sleeve. (Are there *important* norms which are not strictly sanctioned, and are there *unimportant* norms which are strictly enforced?)
* As in other uses of it, the degree of excess expressed by ‘>>’ is left deliberately vague. (To be set, perhaps, by empirical research.)
1. *Norms as Interests*

An alternative approach is to treat the norms to be accepted as interests of the agent. In that case whether the action produced is that required by the relevant norm (or consistent with the norm) is dependent upon the weighting associated with the norm interest. A reasonably obvious way to do is given here.

We shall say that the norm *n* (Normatively) Models an agent *x* iff *x* follows the rule *n* in all circumstances. We will write *v*(*n, x*) for

(*cx*)[*g*(*n*, *cx*)  *n,cx*(*ax*)]

The (Normative) Modelate of the norm *n*, *V*(*n*), is all the agents modeled by *n*. We will write

*V*(*n*) = {*x*: *v*(*n*, *x*)}

Let the finite set *Ix* = {*ix,1*, …, *ix,n*} be the interests of the individual *x*.

Let *Wx* = {*wx,1*, …, *wx,n*} be the set of interest weights of the individual *x*.

We shall say that *x* I-Accepts the norm *n*, and write *k­I*(*x*, *n*) iff

(*j*)[(*ix,j* = ‘*x* *V*(*n*)’) & (((~*g*(*n*, *cx*)  *n,cx*(*ax*))  *Sx*(*C*(*ax*, *cx*), *j*) = 1)  *Sx*(*C*(*ax*, *cx*), *j*) = 0)]

» *One of the interests of the agent is that it is governed by the norm, and the partial satisfaction of this interest is complete in the case that the action is consistent with the norm (or the circumstances aren’t governed by the norm) and null if the action contradicts the norm.*

* The implication of this is that actions consistent with an ‘I-accepted’ norm contribute the interest weight of that norm to the total satisfaction function for that action, whereas actions not consistent with that norm contribute nothing of the interest weight. The question must be whether this properly captures the effect that norm interests (so to speak) have for an agent.
* The notation ‘*x*’ is used to indicate that the interest of the agent *x* does not refer to a variable *x* but to the agent itself. (One cannot substitute into quotations *salva veritate*.)

As previously, for later use we introduce the following terms.

An agent *x* is an I-Acceptant of the norm *n* iff *x* I-accepts the rule *n*.

The I-Acceptate of the norm *n*, *KI*(*n*), is all the agents that I-accept *n*. We will write

*KI*(*n*) = {*x*: *kI*(*x*, *n*)}

* An agent *x* is an I-acceptant of the norm formation *N* iff *x* I-accepts the rules in *N*.
* The I-acceptate of the norm formation *N* is the set of all agents who I-accept the rules in *N*.

*KI*(*N*) = {*x*: (*n* *N*)[*kI*(*x*, *n*)]}

*Norm Comparisons*

The interest approach to norm acceptance described above also provides a model which may be applied to the situation of an agent that is an I-acceptant of norms which may not be univocal. There are many such possible situations, and the full range need not be listed here, but two versions seem most likely to be problematic:

Consider the norms *n1*, *n2*. Let *x*  *K*(*n1*) *K*(*n2*), *cx* is such that *g*(*n1*, *cx*) and *g*(*n2*, *cx*).

1. Suppose *n1*(*ax*). If ~*n2,cx*(*ax*) or *n2,cx*(~*ax*), then we have Norm Conflict in *cx*.
2. Suppose *ax,1* *ax,2*. If *n1,cx*(*ax,1*) & *n2,cx*(*ax,2*), then we have Norm Competition in *cx*.

In such cases, if *ix,j1* = ‘*x*  *KI*(*n1*)’, *ix,j2* = ‘*x*  *KI*(*n2*)’, then the interest weights for the norms are *wx,j1* and *wx,j2* respectively. These weights then feature appropriately in the total satisfaction function to determine the outcome of the action function.

***Externalized Norms***

The RCT perspective on norms is defended by, for example, Coleman (1990, *Foundations of Social Theory*, Cambridge, MA:Belknap.) In this view (p. 243:)

Those subscribing to a norm, or, as I will say, those holding a norm, claim a right to apply sanctions and recognise the right of others holding the norm to do so. Persons whose actions are subject to norms (who themselves may or may not hold the norm) take into account the norms, and the accompanying potential rewards or punishments, not as absolute determinants of their actions, but as elements which affect their decisions about what actions it will be in their interests to carry out.

The RCT view is thus that the observation of norms is purely prudential. We leave the question of the enforcement of norms by the agent for possible later consideration, and deal here with the motivated observation of norms. The consideration of the nature and origin and justification of ‘rights’ need not detain us, as the role that they play here and in other similar presentations is adequately modelled by the relevant expectations.

We say that an agent *x* P-Accepts (P for prudentially) the norm *n*, and write *kP*(*x*, *n*) iff

*Bx­*[ (*cx**ax,*) (*y, qy*) (*cy*) (*ay,hurt**ay,neut*) [

[(*Tx*(*C*(*ay,harm*, *cx*), *Ix*) < *Tx*(*C*(*ay,neut*, *cx,*), *Ix*)) &

(~*By­*[ *g*(*n*, *cx*) & *ax*  *x,bad* *Ay*(*qy*, *cy*) = *ay,neut* ) &

(*By­*[ *g*(*n*, *cx*) & *ax*  *x,bad* *Ay*(*qy*, *cy*) = *ay,harm* )] ]

» *The agent believes that others will modify their actions to his disadvantage if they believe that he has breached a norm.*

* Note the agent-indicator subscripts on the total satisfaction functions and the arguments to them.
* In the case of the externalized norms, to *P-accept* a norm is no more than to accept that there are consequences for breaching it.

As previously, for later use we introduce the following terms.

An agent *x* is a P-Acceptant of the norm *n* iff *x* P-accepts the rule *n*.

The P-Acceptate of the norm *n*, *KP*(*n*), is all the agents that P-accept *n*. We will write

*KP*(*n*) = {*x*: *kP*(*x*, *n*)}

* An agent *x* is a P-acceptant of the norm formation *N* iff *x* P-accepts the rules in *N*.
* The P-acceptate of the norm formation *N* is the set of all agents who P-accept the rules in *N*.

*KP*(*N*) = {*x*: (*n* *N*)[*kP*(*x*, *n*)]}

We define the Governors of the norm *n*, *G*(*n*), to be the set of all agents, *x*, who will enforce *n* and say that *x*  *G*(*n*) iff

(*y, cy, ay*) (*cx, qx* ) (*ax,harm**ax,neut*)

[

(~*Bx­*[ *g*(*n*, *cy*) & *ay*  *y,bad*  *Ax*(*qx*, *cx*) = *ax,neut* ) &

(*Bx­*[ *g*(*n*, *cy*) & *ay*  *y,bad*

(*Tx*(*C*(*ax,harm*, *cx*), *Ix*) << *Tx*(*C*(*ax,neut*, *cx,*), *Ix*)) &

(*Ty*(*C*(*ax,harm*, *cy*), *Iy*) < *Ty*(*C*(*ax,neut*, *cy,*), *Iy*)) 

 *Ax*(*qx*, *cx*) = *ax,harm* )

 ]

» *If an agent believes that some other agent has breached a norm then it will modify its action in such a way that it believes it will disadvantage that other without seriously disadvantaging itself.*

* Note the agent-indicator subscripts on the total satisfaction functions and the arguments to them.
* Note that the P-Acceptate of a norm and its Governors are distinct sets. There are certainly well-known examples where the two have significantly different extensions. The norms associated with the Indian caste system, or those of the Jim Crow South, are two well-studied and serious examples, but examples of less gravity can also be found in most schoolyards.
* In this phrasing it is not claimed that the actions do have the effects claimed, only that it is believed that they do
* In this phrasing, the ‘enforcing’ agent is not required to *intend* to specifically harm the ‘transgressing’ agent, which is probably reasonable, but it *believes* that this may be the case. An obvious modification to the condition would remove even that degree of intentionality.

*Norm Comparison*

Consider again the situations of norm conflict and competition described above. The approach to norm acceptance given here provides a natural model of how an agent could cope with them,

Consider an agent *x* and the twoconflicting or competingnorms *n1* and *n2*.

Suppose that *g*(*n1*, *cx*), *g*(*n2*, *cx*), *kP*(*x*, *n1*), and *kP*(*x*, *n2*).

Suppose that (*kP*(*x*, *n1*) & ~*kP*(*x*, *n2*)) *Ax*(*qx*, *cx*) = *ax,n1* and *ax,n1* & *n2* ⊢ ;

» *If the agent P-accepted n1 but counterfactually did not P-accept n2 then it would do ax,n1 which is disallowed by n2*

and similarly, suppose that (~*k*(*x*, *n1*) & *k*(*x*, *n2*)) *Ax*(*qx*, *cx*) = *ax,n2* and *ax,n2* & *n1* ⊢ .

» *If the agent P-accepted n2 but counterfactually did not P-accept n1 then it would do ax,n2 which is disallowed by n1*

Let *An1, harm* = {*ay, harm*: *y*  *G*(*n2*) & *By­*[ *g*(*n2*, *cx*) & *ax,n1*  *y,bad* ]}

» *The set of all actions taken by the governors of n2 who find ax,n1 to be in breach of n2*

Let *An2, harm* = {*ay, harm*: *y*  *G*(*n1*) & *By­*[ *g*(*n1*, *cx*) & *ax,n2*  *y,bad* ]}

» *The set of all actions taken by the governors of n1 who find ax,n2 to be in breach of n1*

Then:

*Ax*(*qx*, *cx*) = *ax,n1* : *Ex*[*Tx*(*C*(*An1,harm*, *cx*), *Ix*) ] > *Ex*[ *Tx*(*C*(*An2,harm*, *cx,*), *Ix*) ]

 *ax,n2*

» *Act so that you respect the norm whose breaching would result in the most harm from its governors*

* The definitions proposed here can be extended without difficulty to the case of any number of conflicting or competing norms.

***Generalized Norms***

In all that follows the distinction amongst belief, interest, and prudential norms will typically be ignored where it is irrelevant to the effects of interest; thus we will say *x* Accepts a norm, and speak of Acceptants and Acceptates, and write *K*(*n*) and *K*(*N*) with the obvious meanings.