**Agency**

The most basic form of RCT assumes that the agent is first of all an actor, which acts to improve the degree of satisfaction of its interests (broadly understood.) Therefore we begin by considering interests and the measurement of their satisfaction.

In what follows let *x* be an agent and a member of a population *P*.

***Interests and Satisfactions***

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

» *Someone’s ‘interests’ are the set of those events, resources, relations, etc. whose absence or presence are understood to be motivators for that person.*

* *Ix* *I*, the set of all possible interests for any individuals.
* Interests may include ‘needs’ such as food, sex, etc. (Refer to Maslow’s hierarchy for one of many analyses that make the distinction negligible.)
* It is unclear as yet what level of specification is appropriate for sociological analysis. (It is not obvious, for example, whether or not a very specific interest of an individual should be listed as such or subsumed as a part of a more general interest.)
* Where there is no reasonable ambiguity we will write *ii* in place of *ix,i*
* Interests may also be indexed by time.

Let *x* = {*cx,1, …, cx,n*, …} be the set of possible Contexts of *x*.

» *Someone’s ‘context’ is the set of facts about the external environment of that person relevant to the motivators for their actions. These facts may include the number of friends, the employment, the assets held, the social status, etc. The set of their possible contexts is a purely notional class of all the ways that their relevant external environment might be.*

* *x* **, the set of all possible contexts for any individuals.
* Where there is no reasonable ambiguity we will write *cx* in place of *cx,i*
* Contexts may also be indexed by time.
* Each context *cx* is a set of propositional descriptions of relevant facts concerning *x*.
* The context variable refers to the total social context of the agent, for any particular calculation it is likely that the relevant parameters are only a very restricted set of all possible parameters. Approximations and simplifications will certainly make use of this fact.
* Notionally, at least, in any calculation, it is always possible to use the total social context of the entire population, *cP*, which is independent of *x*. Again, however, this is excessive, since the relevant contexts are in almost all cases only those of a very restricted set of all possible agents; generally, of just one in fact, though we will have occasion to expand that concern. *cx* is introduced rather than *cp* in order to indicate that restricted set of agents of interest.
* Where we need to consider the contexts of a set, *X*,of agents we will write *cX* = *x**X cx*

For clarity in the concept, we will say that each context sets the values for *N* Context Parameters, *1*, …, *N*. We can suppose *cx*  *x* is an *N*-dimensional vector (*x,i*)*i=1, N*

* Contexts parameters may also of course be indexed by agent, time, etc.
* The range of each context parameter, *x,i*, is *R*(*x,i*,)
* The assumption that *N* is finite is justified on the grounds that the context is intended to be, at least notionally or potentially, an argument for functions describing actions, or to play the role of a limiting condition on possible claims. That cannot be the case if the function is claimed to have an infinite number of parameters

The degree to which some interest of an agent, *x*, is satisfied is entirely determined by the context of that agent; therefore let us propose the existence of a Partial Satisfaction Function for *x* described as:

*Sx*: *x* *Ix* *ℝ*

» *Anyone has a degree of satisfaction with respect to one of their interests that depends entirely upon the context in which that person currently exists.*

which will generally occur as:

*Sx*(*cx*, *j*) = *sx.j*, the Partial Satisfaction for *x* of the interest *j* *Ix* in the context *cx*.

* Each partial satisfaction function, *Sx*, will be specific to the particular relevant interest, *j*, so it would probably be reasonable to consider *Sx* as *j**Ix Sx,j* for a set of distinct function *Sx,j* if we assumed *Sx,j*(*cx*, *j’*) = 0 wherever *j*  *j’*.
* We may omit most of the variables and indices where there is no ambiguity.
* The *input* to a satisfaction function may be difficult to quantify. For a financial interest this may be easily done by reference to monetary recompense, but other interests have less obvious inputs. What, for example is the proper measure of social interests such as having healthy friendships? And in either case, how do we determine a *degree of satisfaction* with the inputs. These are questions to be addressed empirically: the problem of specification is only significant here if there is *in principle* no way to make sense of it.
* The *output* of a partial satisfaction function is a scalar, positive or negative, but each function will need to be adjusted if arithmetical combinations of satisfactions for different interests are to make sense.
* As several of the arguments of the function may be indexed by time, so are the outputs.

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

» *The degree to which the satisfaction of any interest contributes to overall agent satisfaction will be different for different interests, and different for each agent.*

* Each *wx,j* is the interest weight for the interest *j* of *x*. They are scalars representing the relative value to *x* of the interests involved and function to allow coordination of the values of the outputs of the partial satisfaction functions.
* Where there is no reasonable ambiguity we will write *wi* in place of *wx,i*

We can then define a Total Satisfaction Function for *x* described as:

*Tx*: *x* *Wx**Ix* *ℝ*

» *Anyone has a degree of satisfaction with respect to the totality of their interests that depends entirely upon the context in which that person currently exists. Their total satisfaction is the weighted sum of all their partial satisfactions.*

which will generally occur as

*Tx*(*cx*, *Wx*, *Ix*) = *wx,1Sx*(*cx*, *i1*) + … + *wx,nSx*(*cx*, *in*) as a measure of the Total Satisfaction of *x*

* Generally, we can omit the argument *Wx*.
* *Tx*(*cx*, *Ix*) = *j**ℕ wx,jSx*(*cx*, *ij*) is a briefer way to write the satisfaction function.

***Rational Agency***

The nature of an agent is action. Assuming that the actions of an actor are determined by pre-existing conditions of the world, we propose the existence of a function to relate those pre-existing conditions to the actions taken. A preliminary proposal could describe the Action Function of *x* as:

*A*x: *x*  *x*

» *Actions are determined by the context of an agent.*

where:

1. *x* = {*cx,1, …, cx,n*, …} where each element is a possible context in which *x* acts,
2. *x* = {*ax,1, …, ax,n*, …} is the range of possible actions of *x*.
* *x* **, the set of all possible actions for any individuals.

Each possible action *ax,i* of *x* may result in a modification of the social environment. The social environment typically affects other agents than just *x*. We therefore need to speak of the Context Consequence of action *ax,i* of *x* on *y*. The context consequence is delivered by the Context Consequence Function, described as:

*C*: *x**y*  *y*

» *An action by an agent in the context of an agent will change that context.*

which will generally occur as:

*C*(*ax,i*, *cy,i*) = *cy,i+1*

* *cy,i* is the context of *y* at the time, *i*, that *x* acts.
* *cy,i+1* is the context of *y* that results from the action.
* *C*(*ax,i*, *cx,i*) = *cx,i+1* is the special case of the effect of the action on the agent’s own context
* The same notation can be used unambiguously to refer to the outcomes of a range of *x*’s possible behaviours; thus for the range *x*’= {*ax,1, …, ax,m*}  *x*:

*C*(*x*’, *cy,i*) = {*C*(*ax,1*, *cy,i*)*, …, C*(*ax,m*, *cy,i*)}

= {*cy,i+1,1, …, cy,i+1,m*}

* And we can extend this further to yield:

*C*(*X*’, *cY,i*) = {*C*(*x*’, *cY,i*): *x*  *X, y*  *Y* }, where

* 1. *X*, *Y* are sets of agents
	2. *X*’ is a range of possible actions for each *x* in *X, x*’  *x*
* In what follows we shall consider just the case that *X* = *x* = *y* = *Y*.

The fundamental assumption of RCT, however, is that the rational agent *x* will maximize its total satisfaction. Since this requires that the function be sensitive to the specific interests of the agent, we would need to modify the preliminary claim above concerning the action function; which would thus be described as:

*A*x: *Ix*  *x*  *x*

» *Actions are determined by the context of an actor and its interests.*

which will generally occur as:

*Ax*(*Ix*, *cx,i*) = *ax,i*

where *ax,i* is defined by the relation:

*Tx*(*C*(*ax,i*, *cx,i*), *Ix*) = max{*Tx*(*C*(*ax*, *cx,i*), *Ix*): *ax*  *x*}

 = max{*j**ℕ wx,jSx*(*C*(*ax*, *cx,i*), *ix,j*): *ax*  *x*}

» *The action taken by an agent in any context is that which would lead to a greater satisfaction of its interests in the context that would result than for any other action.*

* It is assumed, reasonably enough for our purposes, that *ax,i* as defined is unique.

***Intentional Agency***

Such is the nature of RCT. As we have said, however, this is widely thought to be an unsatisfactory model of action. We shall be modifying the model of action by assuming instead that the agent in question is an *intentional* agent whose actions are in principle explicable in terms of reasoning from its beliefs directed towards the satisfaction of its desires. As we imagine *x* to be an intentional agent – one that can be considered as a belief-desire-action (BDA) system, we therefore need to introduce operators to refer to the appropriate functions.

For any functions or variables, *V* say, the value that they are estimated to be by *x* is denoted *Ex*[ *V* ].

* Read it as the Estimate of *x* concerning *V*
* The point of including *x* in that notation is that later we will want to be able to account for subjective judgements by *x* of subjective judgements by *y*, etc.
* Where no reasonable likelihood of confusion about the subject exists we may simply write *E*[ *V* ]

For any proposition, *p* say, the claim that *x* Believes that *p* shall be written *Bx*[ *p* ].

Let **B***x* = {*p*: *Bx*[ *p* ]} be the Beliefs of *x*.

* We need to make no commitment at this stage concerning the nature of belief, other than the claim that beliefs are sociologically relevant, in that they are accessible to the mechanisms by which the agent’s actions are determined – in so far as the agent is a BDA system.
* The sociologically relevant beliefs are certainly not restricted to the *explicit* beliefs of an agent, nor even to the beliefs which the agent is ‘aware’ that they have.
* Estimation and belief are related by the equivalence

*Bx*[ *p* ]  *Ex*[ *TV*(*p*)] = 1

» *Someone believes that p, if and only if they think it is true that p.*

with the comments on belief carrying over to estimation. (*TV* is the truth valuation.)

Let **K***x,i =* {*p*: (*cx**x*)[~(*cx* ⊨ *p*) & (*cx,i* ⊨ *p*)]} be the Knowable Context of *x* at *i*. In the case of the agent with perfect knowledge (an effective assumption of RCT,) **K***x,i*  **B***x,i.*

» *The knowable context is all the contingent truths about the world. Whether that counts as knowable in a stronger epistemological sense need not be considered here. The agent with perfect knowledge believes at least all the contingent truths of the world.*

For any functions or variables, *V* say, the value that they would be preferred to be by *x* shall be denoted *Fx*[ *V* ]

* Read it as the Preference of *x* concerning *V*
* Where no reasonable likelihood of confusion about the subject exists we may simply write *F*[ *V* ]

For any proposition, *p* say, the claim that *x* Desires that *p* shall be written *Dx*[ *p* ].

Let **D***x* = {*p*: *Dx*[ *p* ]} be the Desires of *x*.

* We need to make no commitment at this stage concerning the nature of desire, other than the claim that desires are sociologically relevant, in that they are accessible to the mechanisms by which the agent’s actions are determined – in so far as the agent is a BDA system.
* The sociologically relevant desires are certainly not restricted to the *explicit* desires of an agent.
* Preference and desire are related by the equivalence

*D­x*[ *p* ]  *Fx*[ *TV*(*p*)] = 1

» *Someone desires that p, if and only if they want it to be true that p.*

With the comments on desire carrying over to preference

For an intentional agent the degree of satisfaction of interests is now a psychological factor of the agent involved, therefore it’s reasonable to assume that there is a maximum degree of satisfaction that can be achieved for any interest. For convenience we can make the range of each *S* identical to the interval (*0*, *1*).

Let *sx,j* be the state of satisfaction of the interest *j* for *x*. For the context *cx,i*, *sx,j* = *Sx*(*cx,i*, *j*). The agent’s desire to increase the satisfaction of the interest *j* can be written as *D­x*[ *sx,j* = 1 ], however, it is understood that *x* does not desire to increase the satisfaction of each interest independently, but, just as in the case of the RCT model above, to increase the total satisfaction of all its desires. This total satisfaction we can write as *tx*, and for the same reason as for *sx,j* we will assume that the range of *tx* is (*0*, *1*).

We write that *tx* = *i=1,…,n* *wx,isx,ji*

* *i=1,…,n* *wx,i*  1

For the agent *x,* the agent’s desire to increase the satisfaction of its interests can be written as
*D­x*[ *tx* = 1 ], and for our intentional agent *tx* = 1  **D***x*

Let it be noted, however, that the perfectly rational intentional agent does not desire the impossible, but only the best possible. In that case:

*D­x*[ *tx* = max{*Tx*(*cx*, *Ix*): *cx**x*}

» *Satisfaction of one’s interests is entirely a matter of the context that one finds oneself in. One desires that the total satisfaction achieved is the maximum total satisfaction taken over all possible contexts.*

We may also assume that:

*D­x*[ {*cx’*: *Tx*(*cx’*, *Ix*) = max{*Tx*(*cx*, *Ix*): *cx**x*}} ]

» *One desires the context that is necessary to maximally satisfy ones interests.*

and that:

*D­x*[ {*ax’*: *Tx*(*C*(*ax’*, *cx*), *Ix*) = max{*Tx*(*C*(*ax*, *cx*), *Ix*)): *ax**x*}} ]

» *One desires to perform the action that is necessary to create the optimal context.*

* Assume too, as seems reasonable, that the defined set is a singleton.

These sets of propositional attitudes form a distinct set of constraints on the action produced by *x*. They are not elements of the (external) context of *x,* but part of what we might call the ‘internal context,’ the ‘character,’ or rather the Quality of the agent. We need to account for this in the action function which we shall modify thus:

*Ax*: *x*  *x*  *x*

» *Actions are determined by the quality and context of an agent.*

where:

1. *x* = {*qx,1, …, qx,n*, …} where each element is a possible quality of *x* at the point of action

» *Someone’s ‘qualities’ are the set of those characteristics, personality traits, cognitive and emotional dispositions, etc. whose absence or presence are understood to affect the reaction of the agent to an environment.*

1. *qx* is the quality of *x* at the point of action
* Quality is a set of conditions that we call qualities of *x.* These determine the internal state of the agent, where the context conditions determine the external state.
* Both **B***x* and **D***x* are elements of *qx*
* **B***x* and **D***x* may be indexed as **B***x,i* and **D***x,i* to speak of the beliefs and desires of *x* at time *i*
* Other elements of *qx* will be identified as their relevance to any plausible action function is discovered.
* This form of the action function is not novel: it has antecedents dating back at least to Lewin’s equation (1935, *A Dynamic Theory of Personality*, NY:McGraw-Hill / 1936, *Principles of Topological Psychology*, NY:McGraw-Hill) that states that the *B*ehaviour of a person is a *f*unction of their present *P*ersonality traits and their present social *E*nvironment, expressed as

*B* = *f*(*P*, *E*)

The action function for the intentional agent, *x*, will generally occur as:

*Ax*(*qx,i*, *cx,i*) = *ax,i*

 = *ax* where *D­x*[ *ax* ]

 = *ax’* where *Tx*(*C*(*ax’*, *cx,i*), *Ix*) = max{*Tx*(*C*(*ax*, *cx,i*), *Ix*)): *ax**x*}

 = *ax’* where *Tx*(*C*(*ax’*, *cx,i*), *Ix*) = max{(*j=1,…,nwx,jS*x(*C*(*ax*, *cx,i*), *ij*): *ax**x*}

* Since the agent is assumed to be ideal, the actual outcome of this calculation is just as under the earlier RCT assumptions. The weighting is irrelevant to that result.

Note that for the ideal intentional agent, where **K***x,i*  **B***x,i* (and ~(**B***x,i* [⊢](https://en.wikipedia.org/wiki/Turnstile_%28symbol%29) )), the context consequence of an action is rationally derivable from complete accurate knowledge of the action and the existing context, so that:

*C*(*ax,i*, *cy,i*) = *cy,i+1*

= *cy**y* ({*p*: **B***x,i* & *ax,i* [⊢](https://en.wikipedia.org/wiki/Turnstile_%28symbol%29) *p*}  *cy*)

» *The context that follows from an action is described by the set of all statements that rationally follow from the statements of the beliefs of the ideal agent and the description of the action, and which occur in some statement of a possible context. The point of the latter condition is simply to restrict the statements in the context description to the sorts of statements that belong in context descriptions – whatever they might be. Note that the agent is assumed to be a perfect reasoner and to know what the action is.*

Note also that the effect of an action cannot now be limited only to the context consequences. Each possible action *ax,i* of *x* may result in an alteration of *qy,i* , the set of socially relevant characteristics – what we are calling the qualities – of another agent, *y*. We call this the Quality Consequence of action *ax,i* of *x* for *y*. The quality consequence is delivered by the Quality Consequence Function, which has the form:

*Q*: ****  **

which will generally occur as

*Q*(*ax,i*, *qy,i*) = *qy,i+1*,

» *An action by one agent may affect the qualities of some other agent.*

* The same notation can be used unambiguously to refer to the quality consequences for *y* of a range of *x*’s possible behaviours; thus for the range *x*’ = {*ax,1, …, ax,m*}  *x*:

*Q*(*x*’, *qy,0*) = {*Q*(*ax,1*, *qy,i*)*, …, Q*(*ax,m*, *qy,i*)}

= {*qy,i+1,1, …, qy,i+1,m*}

And for an ideal intentional agent again with **K***x,i*  **B***x,i*, the quality consequence of an action is rationally derivable from the complete accurate knowledge of the action and the existing quality, so that:

*Q*(*ax,i*, *qy,i*) = *qy,i+1*

= *qy**y* ({*p*: **B***x,i* & *ax,i* [⊢](https://en.wikipedia.org/wiki/Turnstile_%28symbol%29) *p*}  *qy*)

» *The quality that follows from an action is described by the set of all statements that rationally follow from the statements of the beliefs of the ideal agent and the description of the action, and which occur in some statement of a possible quality. Again, the point of the latter condition is simply to restrict the statements in the quality description to the sorts of statements that belong in quality descriptions – whatever they might be.*

***Imperfect Intentionality***

The perfectly informed and perfectly rational agent is acceptable for a first approximation of intentional agency; however, in order to more closely approximate reality, some adjustments must be made to take account of various forms of imperfection. Some limitations on rationality will be necessary on any agent – for example, no agent has perfect knowledge of their context, or the consequences of their actions; other limitations will be dependent upon the nature of the agent, such as the cognitive limitations investigated by, for example, Kahneman and Tversky (as summarized in *Thinking, Fast and Slow*, 2011, Macmillan.)

Furthermore, given such limitations, the ideal partial satisfaction functions of the agent model will certainly be imperfectly applied. Instead of the function *Sx*(*cx*, *j*) – whatever that might be – we will need to apply the function *Ex*[ *Sx*(*cx*, *j*) ] which is a function that returns the Estimated Partial Satisfaction of interest *j* for agent *x* in context *cx*.

* We note that the estimated partial satisfaction function of the agent model may have little to no relationship to the ideal or actual partial satisfaction function. In the agent model its role may be played by an attempt at a direct model of the reasoning processes, or a set of heuristics intended merely to model the outcome of unknown or uncertainly known processes. There may also be other options and combinations of the various approaches. The function *may* roughly approximate the ideal function over a reasonable range of cases.

We wish to define or describe the Estimated Total Satisfaction function, *Ex*[ *Tx* ]. Reasonable assumptions are that:

* It is related in form to the total satisfaction function, *Tx*, and involves a calculation of the likely effects on the various partial satisfactions.
* It is unlikely that the agent considers all the interests that he may have, or that he properly weighs their contributions. Define the Estimated Interest Weights as:

*Ex*[ *Wx* ] = {*Ex*[ *wx,1* ], …, *Ex*[ *wx,n* ]},

* Assume the distributivity of the *E*[ - ] predicate. The assumption is justified by the supposition that the only reasonable way to determine the estimated values involved is to define them in terms of their function in the estimated total satisfaction function.

Thus:

*Ex*[ *Tx*(*cx*, *Ix*) ] = *Ex*[ *j**ℕ wx,jSx*(*cx*, *j*) ]

= *j**ℕ Ex*[ *wx,jSx*(*cx*, *j*) ]

= *j**ℕ Ex*[ *wx,j* ]*Ex*[ *Sx*(*cx*, *j*) ]

» *The estimated total satisfaction is the sum of the estimated partial satisfactions weighted by their estimated weights.*

For each action it is plausible to model the agent as considering a range of Subjectively Possible Context Consequences, rather than the single consequence that is actually determined by logic and the laws of nature: let this be denoted as *Ex*[ *C*(*ax*, *cy,i*) ] = {*cy,1*, …, *cy,m*}

* Each subjectively possible context consequence, *c*, has an associated Subjective Estimate of Probability, *Ex*[ *Prob*(*c*) ], where *c**Ex*[ *C*(*ax*, *cy,i*) ]*Ex*[ *Prob*(*c*) ] = 1
* To determine the satisfaction potential of an action *a*, the agent *x* will consider the likely satisfactions to be had from each subjectively possible context consequence of the action and weight it by the subjective estimate of probability of that outcome.
Thus, for *a*  *x, j*  *Ix*,

*Ex*[ *Sx*(*C*(*a*, *cx,i*), *j*) ] = *c**E*[*C*(*a,cx,i*)] *Ex*[ *Prob*(*c*) ]*Ex*[ *Sx*(*c*, *j*) ]

* We thus require the further modification of the estimated total satisfaction function for the action *a* of *x*:

*Ex*[ *Tx*(*C*(*a*, *cx,i*), *Ix*) ] = *j**ℕ Ex*[ *wx,j* ]*Ex*[ *Sx*(*C*(*a*, *cx,i*), *j*) ]

 = *j**ℕ Ex*[ *wx,j* ]*c**E*[*C*(*a,cx,i*)]*Ex*[ *Prob*(*c*) ]*Ex*[ *Sx*(*c*, *j*) ]

It is certain that the agent will not consider all possible actions. The set of Subjectively Possible Actions that the agent considers live options will be denoted *E*[ *x* ] = {*ax,1*, …, *ax,m*}

Granted these forms of limited rationality, the action produced by the agent will maximize the estimated total satisfaction of all subjectively possible actions, yielding an output action *ax,i*, such that:

*Ex*[ *Tx*(*C*(*ax,i*, *cx,i*), *Ix*) ]

= max{*Ex*[ *Tx*(*C*(*ax*, *cx,i*), *Ix*) ]: *ax*  *Ex*[ *x* ]}

 = max{*j**ℕ Ex*[ *wx,j* ]*c**E*[*C*(*ax,cx,i*)]*Ex*[ *Prob*(*c*) ]*Ex*[ *Sx*(*c*, *j*) ]: *ax*  *E*[ *x* ]}

* I take this to be true almost by definition of action. Argument will arise over the nature of the *Tx* function or the method of determining *Ex*[ *Prob*(-) ].