3

Socio-Cognitive Model of Trust: Quantitative Aspects

So far in this book, we have analyzed trust from the *qualitative* point of view: we have carefully discussed the cognitive ingredients of trust, their relationships and formalizations. Before going on (describing trust dynamics, trust sources, trust generalization, etc.) we have to evaluate, understand, describe and formalize the *quantitative* nature of the trust concept.

It is true that when *Xania* decides to trust *Yody*, she has considered the different aspects of trustworthiness, like ability, willingness, context, and so on, and the various reasons and causes these aspects are based on. But, at the same time, she has also evaluated their amount: if the weight of each element is enough, if the quantity of their complete composition (also considering potential overlapping and interferences) can be evaluated as sufficient for trusting *Yody*. Every day we participate in discussions where judgments like: 'John is really trustworthy, because he knows his work and is very competent and serious with the customers', are expressed, where quantitative evaluations are expressed in an approximate, colloquial way: what does it mean 'really', 'very'? How much does know about John his work? How much is serious? How much are we sure about this? In fact, directly or indirectly, there is always a judgment of quantification over these properties. And we always test these quantifications: 'How much do you trust him?', 'Is your trust in him sufficient?', 'Is he so trustworthy?' 'Are you sure?', and so on.

So although the qualitative analysis of the trust components is fundamental for getting the real sense of trust concept, the quantification of its ingredients and an adequate composition of them will permit the results of its application to be effectively evaluated and simulated.

3.1 Degrees of Trust: a Principled Quantification of Trust

The idea that trust is measurable is usual (in common sense, in social sciences, in AI (Snijders, 1996), (Marsh, 1994)). In fact, in the majority of the approaches to the trust study the quantification aspect emerges and prevails over the qualitative and more analytic aspects (that are considered less relevant and sometimes useless). Because of this, in these approaches

Trust Theory: A Socio-Cognitive and Computational Model Cristiano Castelfranchi and Rino Falcone
 ISBN: 978-0-470-02875-9

no real definition and cognitive characterization of trust is given; so, the quantification of trust is quite *ad hoc* and arbitrary, and the introduction of this notion or predicate results in being semantically empty.[1]

On the contrary, in our studies we try to understand and define the relationships between the cognitive definition of trust, its mental ingredients, and, on the one hand, its value. On the other hand, its social functions and its affective aspects (Chapter 5). More precisely the latter are based on the former.

In this chapter we will show our efforts to ground the degree of trust of X in Y in the cognitive components of X's mental state of trust.[2] In particular, given our belief and evaluation based model, we predict and claim that *the degree of trust is a function:*

- *on the one hand, of the estimated degree of the ascribed 'quality' of Y on which the positive expectation is based;*
- *on the other hand, it is a function of the subjective certainty of the pertinent beliefs.*

Let us be more specific: the first component describes the quantitative level of Y's quality under analysis: for example, if X is evaluating Y's *ability* (about a given task τ) she has to select among different discrete or continuous values the one (or ones) she considers the more adequate to attribute to Y. These values could be either directly numerical or described by linguistic categories referable to a set of numerical attributions ('very good', 'good', 'sufficient', 'poor', just to give some examples). X could have, for example, a main (prevalent) belief that Y is either with ability 0.7 or 0.8 (in the scale (0,1)) and a secondary (less relevant) belief that Y is not so able (ability included in the 0.2-0.3 interval). See Figure 3.1 as an example.

At the same time, X also has a *meta-belief,*[3] about the subjective certainty of these beliefs (the second component indicated above): how much is X sure of her evalutative beliefs about Y's quality?

These meta-beliefs in fact translate the strength of the reasons that produced the first-level beliefs. Are they based on a consistent (or just superficial) set of experiences, reasoning, facts, deductions, a priori judgments, and so on? There is, of course, a correlation between the construction of the first kind of belief and the building of the second kind, but we can distinguish (and it is useful to do this for analytical reasons) between the different semantic and functional roles of the two categories. In any case, for simplicity, in the following part of the chapter we will consider the integration of beliefs and meta-beliefs.

We will use the degree of trust to formalize a rational basis for the decision of relying and betting on Y. We will also consider – for the 'decision' to trust – the quantitative aspect of another basic cognitive ingredient: *the value or importance or utility of the goal* g_X.

As we said, trust always implies risks, and frequently 'perceived' (evaluated) risks (Section 2.8.2). So we will also introduce *the evaluation of the risk* (depending on the potential

---

[1] As reported also in other parts of this book Williamson (Williamson, 1993) claims that 'trust' is an empty and superfluous notion – used by sociologists jn a rhetorical way – since it is simply reducible to subjective probability/risk (Chapter 8).

[2] Precisely the 'affective' trust in part represents an exception to that; since its 'degree' is not based on arguable 'reasons' but it is due to the 'intensity' of the feeling or of the evocated 'somatic markers'.

[3] An 'explicit' belief about other beliefs, or an 'implicit' one; that is, just some index of the belief strength or certainty; like the one we have introduced in the analysis of 'expectations' and of their 'strength' (Chapter 2).

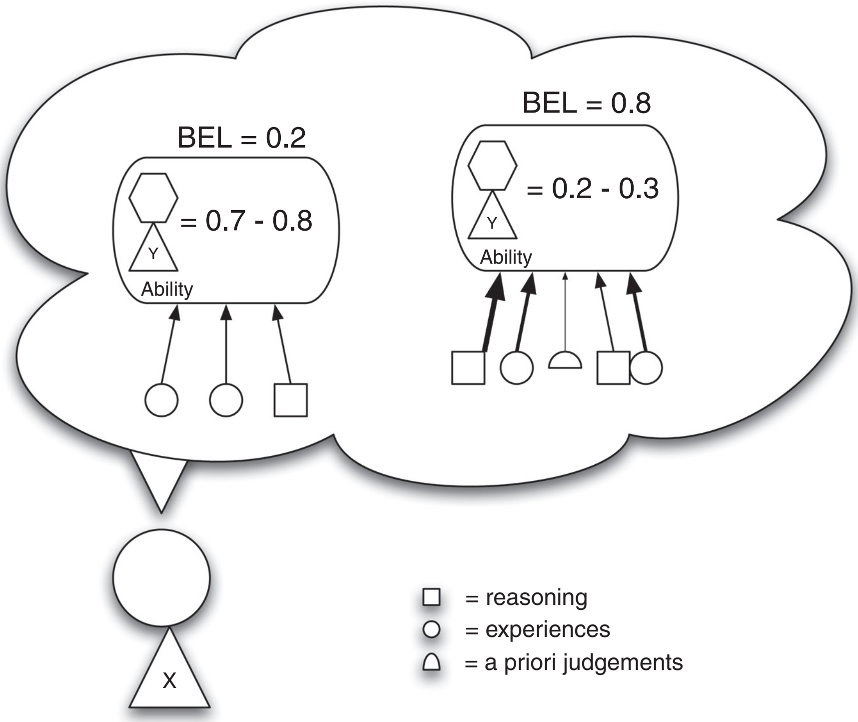

Figure 3.1 Quantifying *X*'s Beliefs about *Y*'s ability

failures of the various components of trust). Finally, we will introduce different quantitative thresholds linked with the different quantitative dimensions included in our cognitive model.

In sum, *the quantitative dimensions of trust are based on the quantitative dimensions of its cognitive constituents*. For us, trust is not an arbitrary index just with an operational importance, without a real content, or a mere statistical result; but it is derived from the quantitative dimensions of the pertinent beliefs.

3.2 Relationships between Trust in Beliefs and Trust in Action and Delegation

The solution we propose is not an *ad hoc* solution, just to ground some degree of trust. It instanciates a general claim. Pears (Pears, 1971) points out the relation between the *level of confidence* in a belief and the likelihood of a person *taking action based on the belief*: 'Think of the person who makes a true statement based on adequate reasons, but does not feel confident that it is true. Obviously, *he is much less likely to act on it, and, in the extreme case of lack of confidence, would not act on it*' (p. 15) (We stressed the terms clearly related to theory of trust).

'It is commonly accepted that people behave in accordance with their knowledge' (notice that this is precisely our definition of a *cognitive agent* but it would be better to use the word 'beliefs' instead of 'knowledge'). '*The more certain the knowledge then the more likely, more rapid and more reliable is the response*. If a person strongly believes something to be correct which is, in fact, incorrect, then the performance of the tasks which rely on this erroneous belief or misinformation will likewise be in error – even though the response may be executed rapidly and with confidence.' (p. 8). (Hunt and Hassmen, 1997). [4]

Thus under our foundation of the degree of trust there is a general principle: Agents act depending on what they believe, i.e. *relying on* their beliefs. And they act on the basis of the degree of reliability and certainty they attribute to their beliefs. In other words, trust/confidence in an action or plan (reasons to choose it and expectations of success) is grounded on and derives from trust/confidence in the related beliefs.

The case of trust in delegated tools or agents is just a consequence of this general principle in cognitive agents. Also, beliefs are something one bets and risks on, when one decides to base one's action on them, although, frequently, without an explicit deliberation about this, but with a procedural and implicit assumption about their reliability. Chosen actions too are something one bets, relies, counts on and depends upon. We trust our beliefs, we trust our actions, we trust delegated tools and agents. In an uncertain world any single action would be impossible without some form of trust (Luhmann, 1990).

3.3 A Belief-Based *Degree of Trust*

Let's call the degree of trust of X in Y about τ:

$$DoT_{XY\tau} \tag{3.1}$$

with $0 \leq DoT_{XY\tau} \leq 1$, where $DoT_{XY\tau}=0$ means absolutely no trust, and $DoT_{XY\tau}=1$ means full trust (in fact, a sort of *faith*): these two values are in fact two asymptotic limits (they are contradictory to the definition of trust as always including some risk:[5] although, strictly subjectively speaking, the risk might be ignored).

As described in Section 3.1, we can distinguish between evaluative beliefs (about the qualities of the trustee or its contextual environment) and meta-beliefs (how much the trustor is sure about that evaluative belief). Suppose X is considering Y's ability (Ab_Y, with $0 \leq Ab_Y \leq 1$) about the task τ. For different reasons (direct experiences, reasoning about categories in which Y is included, and so on) X could have several values to attribute to Ab_Y ($Ab_{1Y}=0.2$, $Ab_{2Y}=0.4$, $Ab_{3Y}=0.8$). Each of these possibilities has different strengths (suppose $Bel_X(Ab_{1Y})=0.7$, $Bel_X(Ab_{2Y})=0.5$, $Bel_X(Ab_{3Y})=0.4$, respectively) (see Figure 3.2).

Imagine, for example (see Figure 3.2), the case in which *Xania* observed *Yody* carrying out a specific task (many years ago: so the strength of the source is not that high (0.40)) performing rather well (0.80), in addition, someone not so reliable as source (0.50) informed her about a

[4] This correct view, is just incomplete; it ignores the dialectic, circular, relationships between action and beliefs: a successful action – based on certain assumptions – automatically and unconsciously reinforces, 'confirms' those beliefs; an unsuccessful action, a failure, arouses our attention (surprise) about the implicit assumptions, and casts some doubt over some of the grounding beliefs.

[5] In the case of $DoT_{XY\tau}=1$ we would have no risks, the full certainty of Y's success. In the case of $DoT_{XY\tau}=0$ we would have absolute certainty of Y's failure (analogously no risks, that subsume uncertainty).

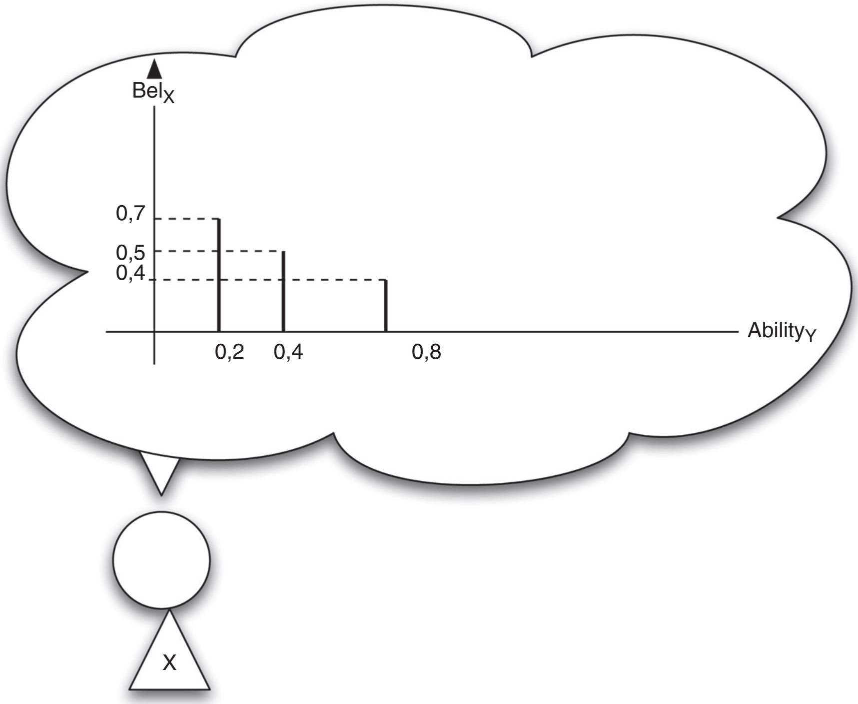

Figure 3.2 Relationships among different evalutative Beliefs and their strenghts

modest ability of *Yody* (0.40) and at the same time *Xania* believes (with high certainty (0.70)) *Yody* belongs to a category which, generally (for its own features), performs very badly (0.20) on those kinds of task.

At this stage of the analysis it is also possible to have concurrent hypotheses with similar values (the sum should not necessarily be equal to *1*: again, they are not coherently integrated but just singularly evaluated). The meta-belief (we can also consider it the strength of this belief) on *Y*'s ability has to consider the different beliefs about this quality[6] and, on the basis of a set of constraints (see later), define the degree of credibility *DoC* of the specific quality.

In general, we could write that:

$$DoC_X(Qual\text{-}i_{(s1,\ldots sn),Y}(\tau)) = F_{X,Y,\tau}(Bel_X(Str_1 Qual\text{-}i_{s1Y}(\tau)), \\ Bel_X(Str_2 Qual\text{-}i_{s2Y}(\tau)), \ldots, Bel_X(Str_n Qual\text{-}i_{snY}(\tau))) \qquad (3.2)$$

In other words, the degree of credibility for *X* (DoC_X) about the *i*-th *Quality of Y* on the task τ ($Qual_{iY}(\tau)$) on the basis of the *n* belief sources ($S_1, \ldots, S_n$) is *a function (depending on X, Y*

[6] Here we are considering *Y*'s ability as a quality of *Y*. We have used the term quality in Chapter 2 at a more detailed level: the set of features characterizing an entity. In fact *Y*'s ability might be considered as a meta-quality (composed by a set of more specific qualities).

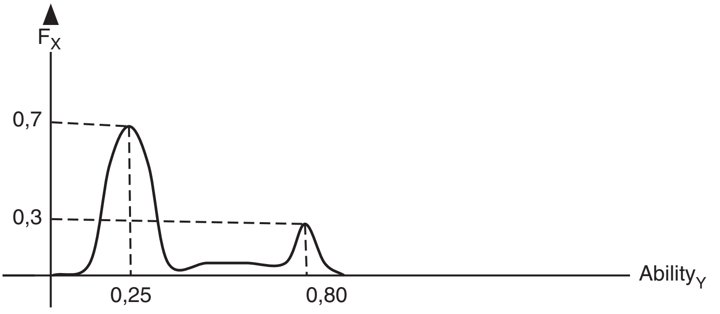

Figure 3.3 Probabilities for two different values of Y's ability to perform the task τ (as output of the function F)

and τ) of the different strengths of the belief sources (in X's point of view: Str_1, Str_2, . . ., Str_n) that weigh the different contents (Qual-$i_{s1Y}(\tau)$, Qual-$i_{s2Y}(\tau)$, . . ., Qual-$i_{snY}(\tau)$).

$F_{X,Y,\tau}$ is a *selective and normalizing function* that associates couples *quality-values/ strengths-of-beliefs* with a probability curve (with all the constraints the probability model introduces).

In clearer terms, $F_{X,Y,\tau}$ should produce a matrix with n rows and two columns, where the number of rows corresponds with the *number* of *quality-values* that have to be taken into consideration by the selective process (not necessarily all the values reported by the belief sources[7]) while the two columns represent the *contents* of the *quality-values* and their *normalized probabilities* respectively (the sum of the probabilities in the second column must be equal to *1*).

As an example we can show how, starting from Figure 3.2, $F_{X,Y,\tau}$ by selection (for example the values *0.2* and *0.4* are collapsed in *0.25*) and by normalization (producing *0.7* and *0.3* as probabilities values) gives the output of the matrix:

$$\begin{pmatrix} 0.25 & 0.7 \\ 0.80 & 0.3 \end{pmatrix}$$

or, to see this in a clear visual form see the graph in Figure 3.3

In the following, for the sake of simplicity, $DoC_X(\textit{Qual-}i_{(s1,\ldots sn),Y}(\tau))$ will give as a result the more probable value of that quality weighted on the basis of its strength and of the probability of the other results and strengths. In other words, *we reduce the function F to a single number averaging the complexity of the different values, sources and strengths.*

Given that we postulate that the degree of trust basically is a function of the 'strength' of the trusting beliefs, i.e. of their *credibility* (expressing both the subjective probability of the fact and trust in the belief): the stronger X's belief in Y's competence and performance, the stronger X's trust in Y.

[7] In some cases maybe there is an integration of some of these values.

Assuming that the various credibility degrees are independent from each other, we can say (for simplicity we consider $p{=}g_X$):

$$DoT_{XY_\tau} = C_{opp} DoC_X[Opp_Y(\alpha,p)]^* C_{ab}\ DoC_X[Ability_Y(\alpha)]^* \times C_{will}\ DoC_X[WillDo_Y(\alpha,p)] \tag{3.3}$$

where:

- C_{opp}, C_{ab}, and C_{will} are constant values and represent the weights of the different credibility terms[8]: they take into account the variable *relevance* or *importance* of the different components of Y's trustworthiness. Depending on the kind of task, on X's personality, etc, Y's *competence* or Y's *reliability* do not have equal impact on his global trustworthiness for task τ, and on X's decision.

If, for example, τ is quite a technical task (like repairing an engine, or a surgical intervention) Y's competence is more important, and its *weight* in the evaluation and decision is more determinant; if τ is not technically demanding but its deadline is very important then Y's punctuality or reliability is more relevant than his competence or skills.

- $DoC_X[Opp_Y(\alpha,p)]$, is the degree of credibility of X's beliefs (for X herself) about Y's opportunity of performing α to realize p; in more simple words, it takes into account all the contextual factors in which Y is considered to act.
- $DoC_X[Ability_Y(\alpha)]$, the degree of credibility of X's beliefs (for X herself) about Y's ability/ competence to perform α;
- $DoC_X[WillDo_Y(\alpha,p)]$, the degree of credibility of X's beliefs (for X herself) about Y's actual performance.

In a case in which Y is a cognitive agent, the last degree ($DoC_X[WillDo_Y(\alpha,p)]$) will become:

$$DoC_X[WillDo_Y(\alpha,p)] = DoC_X[Intend_Y(\alpha,p)]^* DoC_X[Persist_Y(\alpha,p)]$$

and can be interpreted as the degree of credibility of X's beliefs (for X herself) about Y's willingness to actually perform α to realize p; where the willingness can be split in the composition of intention and persistence.

Finally, of course:

$$0 \le DoC_X[Opp_Y(\alpha,\ p)] \le 1;\ 0 \le DoC_X[Ability_Y(\alpha)] \le 1;$$
$$0 \le DoC_X[WillDo_Y(\alpha,p)] \le 1.$$

3.4 To Trust or Not to Trust: Degrees of Trust and Decision to Trust

In this paragraph we analyze the complex process of taking a real (reason-based) decision about trusting or not, on the basis of the mental ingredients described in Chapter 2 and of their quantitative values.

[8] In fact the role of these factors would be more complex than simple constant values, they should represent the set of non linear phenomena like saturation effects, possible interference among the different credibility degrees, and so on.

Resuming the trustor's mental state we have two main subsets:

(a) A set of mental states ($MS\text{-}CT_{X,Y}$) -called *Core Trust*- with these components:
   - a set of *X's goals* and, in particular, one specific of them (g_X) in order to trust Y;
   - a set of *X's competence beliefs* ($B\text{-}Com_{X,Y}$) on Y about τ;
   - a set of *X's disposition beliefs* ($B\text{-}Dis_{X,Y}$) on Y about τ and
   - a set of *X's practical opportunities beliefs* ($B\text{-}PrOp_{X,Y}$) on Y about τ at that given moment (time) and site (space).

(b) A set of mental states ($MS\text{-}REL_{X,Y}$) -called *Reliance*- that must be added to the 'core trust' ones and that are strictly linked with the decision to trust; in particular:
   - a set of *X's dependence beliefs* ($B\text{-}Dep_{X,Y}$) (it is needed or it is better to delegate than not delegate to Y ((Sichman *et al.*, 1994), (Jennings, 1993)) and
   - a set of *X's preference beliefs* ($B\text{-}Pref_{X,Y}$) for delegating to Y (in fact, although this notion is related to the dependence notion, we like to mark it).

We can imagine that each one of the above listed beliefs will have a specific value.

In order that X trusts Y about τ, and thus delegates that task to Y, it is not only necessary that the $DoT_{XY\tau}$ exceeds a given threshold (depending on X, Y and the task τ), but also that *it constitutes the best solution* (compared with other possible and practicable solutions).

In any circumstance, an agent X endowed with a given goal, has three main choices:[9]

i) to try to achieve the goal by itself;
ii) to delegate the achievement of that goal to another agent Y;
iii) to do nothing (relative to this goal).

So we should consider the following abstract scenario (Figure 3.4)
where we call:

$U(X)$, the agent X's utility function, and specifically:

$U(X)_{p^+}$, the utility of X's success performance (directly realized by agent X);

$U(X)_{p^-}$, the utility of X's failure performance (directly realized by agent X);

$U(X)_{d^+}$, the utility of a successful delegation (utility due to the success of the delegated action to Y);

$U(X)_{d^-}$ the utility of a failure delegation (damage due to the failure of the delegated action to Y);

$U(X)_0$ the utility of doing nothing.

[9] The choice of collaborating on a given goal implies the agreed delegation of a subgoal (of a subpart of the main goal) to Y: that means to apply the choice number (ii).

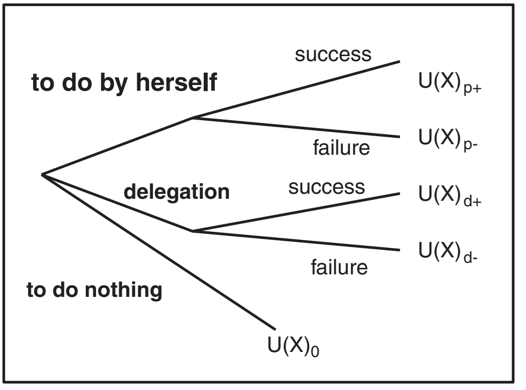

Figure 3.4 Decision Tree for Trust-based Delegation. (Reproduced with kind permission of Springer Science+Business Media © 2001)

One should also consider that trust (the attitude and/or the act) may have a value *per se*, independently on the achieved results of the 'delegated' task ((McLeod, 2006) *Stanford Encyclopedia*). This value is taken into account in the decision. For example, if to have/show/put trust in *Y* is a positive thing (for the subject or for the social environment *X* cares about) this 'value' (the satisfaction of this additional goal) should be included among the results (outcomes) of the decision to trust *Y*; and it might be determinant for the choice, even winning against worries and doubts. Vice versa, if trust is a negative fact (for example, a sign of naivety, of weakness of character, etc., of stupidity) this effect too will be taken into account in the decision.

Among the positive results of the act of trusting (successful or unsuccessful; see Figure 3.4) *X* will put the value of trusting *in se* and *per se*; in the opposite case, among the positive results of the act of trusting (successful or unsuccessful) *X* will put the negative value (cost, harm) of trusting *in se* and *per se*.

However, for the sake of simplicity, we will consider the following scenario (Figure 3.5): In the simplified scenario, in order to delegate we must have (using the Expected Utility Theory approach by Bernoulli and von Neumann and Morgenstern (von Neumann and Morgenstern, 1944)):

$$DoT_{XY\tau}\,{}^*U(X)_{d^+} + (1\text{-}DoT_{XY\tau})U(X)_{d^-} > DoT_{XX\tau}\,{}^*U(X)_{p^+} + (1\text{-}DoT_{XX\tau})\,U(X)_{p^-} \tag{3.4}$$

where $DoT_{XX\tau}$ is the *selftrust* of *X* about τ.

Analyzing more carefully the different kinds of utilities we can say that:

$U(X)_{p^+} = Value(g) + Cost\ [Performance(X\ \tau)],$

$U(X)_{p^-} = Cost\ [Performance(X\ \tau)] + Additional\ Damage\ for\ failure$

$U(X)_{d^+} = Value(g) + Cost\ [Delegation(X\ Y\ \tau)],$

$U(X)_{d^-} = Cost\ [Delegation(X\ Y\ \tau)] + Additional\ Damage\ for\ failure$

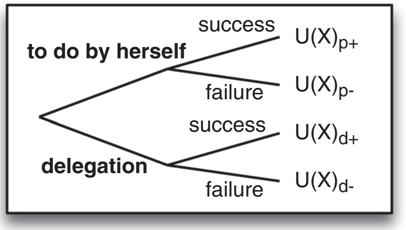

Figure 3.5 Simplified Decision Tree for Trust-based Delegation

where it is supposed that it is possible to attribute a quantitative value (importance) to the goals and where the values of the actions (delegation and performance) are supposed to be negative (having costs: energies and resources committed in the enterprise).

We have called *Additional Damage for Failure* the negative results of both the failures (direct performance and delegation): they represent not only the frustration of the missed achievement of the goal but also the potential additional damages coming from these failures (in terms of compromising other goals, interests, resources, and so on): in particular, in the case of trusting *Y* there are all the risks of reduced control in the situation following the delegation act.

From the formula (3.4) we obtain:

$$DoT_{XY\tau} > DoT_{XX\tau}{}^{*}A + B \tag{3.5}$$

where:

$A = (U(X)_{p^+} - U(X)_{p-}) / (U(X)_{d^+} - U(X)_{d^-})$
$B = (U(X)_{p^-} - U(X)_{d^-}) / (U(X)_{d^+} - U(X)_{d^-})$

Let us consider now, the two terms *A* and *B* separately.

As for term *A*, (considering *B=0*, that means that the trustor has the same damage in terms of cost in the case of the failure of the direct performance and in the case of the failure of action delegated to *Y*) if:

- *A=1* (the differences between success utility and failure utility in the case of *X*'s direct performance and in the case of delegation to *Y* are the same):

then the formula (3.5) becames, $DoT_{XY\tau} > DoT_{XX\tau}$ in practice, *in this case (B=0 and A=1) it is necessary more trust in Y than in herself for X's delegation.*

This (A=1, B=0) is a clear and rational result: in the case in which:

1) *the utility's difference between success and failure is the same (in performing the task by themselves or delegating the task to the trustee)(A=1); and*
2) *the failure of X's performance has the same damage of the failure of the Y's delegated performance; then*
   X has to trust Y more than herself for delegating to him that task: for X is rational to delegate to Y for any value greater than zero between these two trust values ($DoT_{XY\tau}$ - $DoT_{XX\tau}$ >0).

- case $A>1$:

$$U(X)_{p^+} - U(X)_{p^-} > U(X)_{d^+} - U(X)_{d^-}$$

then $A^* DoT_{XX\tau} > DoT_{XX\tau}$

i.e. if the difference between the utility of the success and the utility of the failure in delegation is smaller than the difference between the utility of the success and the utility of the failure in the direct performance of X, then (for the term A) in order to delegate return *the trust of X in Y must be bigger than the selftrust of X* (about τ).

More precisely, suppose two subcases:

(i) In the case of $DoT_{XY\tau} > (DoT_{XX\tau} * a + B)$ *(with $B=0$ and $1<a<A$, $A>1$)* even if the X's degree of trust in Y is greater than the selftrust, *it is not sufficient for delegation*.
(ii) In the case of $DoT_{XY\tau} > (DoT_{XX\tau} * a + B)$ *(with $B=0$ and $a \geq A$, $A>1$), the X's degree of trust in Y is sufficient for delegating*.

In other words, for delegating X's trust in Y has to be greater thnt X's selftrust at least of the value given from A.

Also this ($A>1$, $B=0$) is a clear and rational result: in the case in which:

1) *the difference between the utility of the success and the utility of the failure in direct performance of X is greater than the difference between the utility of the success and the utility of the failure in delegation to Y, and*
2) *the failure of X's performance has the same damage of the failure of Y's delegated performance; then*
X has to trust Y more than herself for delegating to him that task: the difference in the trust values has to be at least of a factor given from A.

Vice versa, if

- $A<1$:

$$U(X)_{p^+} - U(X)_{p^-} < U(X)_{d^+} - U(X)_{d^-}$$

then $A^* DoT_{XX\tau} < DoT_{XX\tau}$

i.e., if the difference between the utility of the success and the utility of the failure in delegation is bigger than the difference between the utility of the success and the utility of the failure in X's direct performance, then (given $B=0$) in order to delegate

the trust of X in Y should be smaller (of A factor) than the X's selftrust (about τ).

Even in this case we have to distinguish two alternative subcases:

(i) In the case of $DoT_{XY\tau} > (DoT_{XX\tau} * a + B)$ *(with $B=0$ and $0<a \leq A$, $A<1$) X's degree of trust in Y is sufficient for delegating*.

(ii) In the case of $DoT_{XY\tau}$ *>*($DoT_{XX\tau}$ * *a* + *B*) *(with B=0 and* A<a<1, *A<1) X's degree of trust in Y is not sufficient for delegation.*

Note that even if in both the cases *X's trust in Y is smaller than X's selftrust, only in one of them is* it possible to delegate: so we can say that (for utilities reasons) *it is possible to delegate to agents which X trusts less than herself*.

Considering now also the term *B* ($B \neq 0$),

- If $U(X)_{p^-} - U(X)_{d^-} > 0$, then a positive term is added to the A: $A + B > A$, i.e., *if the utility of the failure in case of X's direct performance is bigger than the utility of the failure in case of delegation, then - in order to delegate - the trust of X in Y about τ must be greater than in the case in which the right part of* **(3.5)** *is constituted by A alone*.
  Vice versa,
- If $U(X)_{p^-} - U(X)_{d^-} < 0$, then $A + B < A$, i.e., *if the utility of the failure in the case of non-delegating is smaller than the utility of the failure in the case of delegation, then – in order to delegate – the trust of X in Y about τ must be smaller than in the case in which the right part of the formula* **(3.5)** *is constituted by just A alone*.[10]

Since $DoT_{XY\tau} \leq 1$, from the formula **(3.5)** we can obtain (starting from $1 > DoT_{XX\tau} * A + B$):

$$DoT_{XX\tau} < (U(X)_{d^+} - U(X)_{p^-})/(U(X)_{p^+} - U(X)_{p^-}) \tag{3.6}$$

From the formula **(3.6)** we have two consequences in the dynamics of trust; to delegate *X* to *Y* the task τ, as the selftrust ($DoT_{XX\tau}$) grows either:

1) the difference between the utility of the success in delegation and the utility of the failure in the direct performance increases; or
2) it reduces the difference between the utility of the success and of the failure in direct performance.

Because $DoT_{XX\tau} \geq 0$, from **(3.6)** we obtain (starting from $0 < (U(X)_{d^+} - U(X)_{p^-})/(U(X)_{p^+} - U(X)_{p^-})$:

$$U(X)_{d^+} > U(X)_{p^-} \tag{3.7}$$

(consider that for definition we have $U(X)_{p^+} > U(X)_{p^-}$).

In practice, for delegating, a necessary (but not sufficient) condition is that the utility of the success in delegation is greater than the utility of the failure in direct performance *(as intuitively rational).*

Let us conclude this section by underlining the fact that (in our model and in real life) *we do not necessarily delegate to the most trustworthy agent*; we do not necessarily choose the alternative where trust is greater. We might prefer to choose a partner or to rely on a device that is not the most reliable one, simply because there are other parameters involved in our

[10] Both for *A* and *B* there is a normalization factor $(U(X)_d{}^+ - U(X)_d{}^-)$: the more its value increases, the more the importance of the terms is reduced.

decision to delegate (to 'trust', as action): costs, risks, utility and so on. For example, the most competent and trustworthy doctor might be the most expensive or not immediately available (because very busy). In this case, we could delegate a less competent and cheaper one.

Another important thing to be underlined is that the complete scenario of Figure 3.4, with all its branches and precise pros and cons of each alternative, is an ideal situation just rarely effectively evaluated (by humans) in real life (and often also in artificially simulated scenarios): it is more a normative model of a trust decision. However, increasing the importance of the goal to be achieved, the risks of potential damages, and the time for the decision, the choice of enquiring the potential alternative branches becomes a necessity beyond the available and achievable information: it is the paradigmatic scenario upon which trust reasoning must be based.

3.5 Positive Trust is not Enough: a Variable Threshold for Risk Acceptance/Avoidance

As we saw, *the decision to trust is based on some positive trust*, i.e. on some evaluation and expectation about the capability and willingness of the trustee and the probability of success. And on the necessity/opportunity/preference of this delegation act.

First of all, those beliefs can be well justified, warranted and based on reasons. This represents the 'rational' (reasons based) part of the trust in *Y*. But they can also be unwarranted, not based on evidence, even quite irrational, or intuitive (based on sensations or feelings), faithful. We call this part of the trust in *Y*: 'faith'.[11]

Notice that irrationality in trust decision can derive from these unjustified beliefs, i.e. on the ratio of mere faith.

Second, *positive trust is not enough* to account for the decision to trust/delegate. We do not distinguish in this book the different role or impact of the rational and irrational part of our trust or positive expectations about *Y*'s action: the entire positive trust (reason-based + faithful) is necessary and contributes to the *Degree of Trust*: its sum should be greater than discouraging factors. We do not go deeply in this distinction (a part of the problem of rational Vs irrational trust) also because we are interested here in the additional fact that this (grounded or ungrounded) positive expectation can not be enough to explain the *decision/act* of trusting. In fact, another aspect is necessarily involved in this decision: the decision to trust/delegate necessarily implies *the acceptance of some risk*. A trusting agent is a risk-acceptant agent, either consciously or unconsciously. Trust is never certainty: always it retains some uncertainty (ignorance)[12] and some probability of failure, and the agent must accept this and be willing to run such a risk (see Chapter 2) with both positive and negative expectations, and the fact that they can remain just 'implicit' or 'potential'.

Thus a fundamental component of our decision to trust *Y*, is our acceptance and felt exposition to a risk. Risk is represented in previous quantification of *DoT* and in the criteria

[11] To be more precise, non-rational blind trust is close to faith. Faith is more than trust without evidence, it is trust without the need for and the search for evidence, and even against evidence.

[12] We do not want to introduce here a more sophisticated model where 'ignorance' and 'uncertainty' are explicitly represented and distinct from probability; like the Dempster & Shafer model ((Dempster, 1968), (Shafer, 1976)) and the theory of 'plausibility'. We use (less formally) this more sophisticated model in other chapters to explain trust 'optimism' and other aspects.

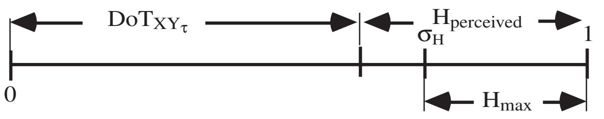

Figure 3.6 Degree of Trust and Hazard Threshold: the case of the Delegation Branch. (Reproduced with kind permission of Springer Science+Business Media © 2001)

for decision. However, we believe that this is not enough. A specific risk policy seems necessary for the trust decision and bet; and we should aim to capture this aspect explicitly.

The equation (3.4) – that basically follows classical decision theory – introduces the degree of trust instead of a simple probability factor. In this way, it permits one to evaluate when to delegate rather than to do it herself in a rigid, rational way. The importance of this equation is to establish what decision branch is the best on the basis of both the relative (success and failure) utilities for each branch and the probability (trust based) of each of them. In this equation no factor can play a role independently from the others. Unfortunately, in several situations and contexts, not just for the human decision makers but – we think – also for good artificial decision makers, it is important to consider the absolute values of some parameter independently from the values of the others. This fact suggests that some saturation-based mechanism, or threshold, by which to influence the decision, needs to be introduced.

For example, it is possible that the value of the damage *per se* (in case of failure) is too high to choose a given decision branch, and this is independent either from the probability of the failure (even if it is very low) or from the possible payoff (even if it is very high). In other words, that danger might seem to the agent an intolerable risk. In this paragraph we analyze (just in a qualitative way) different possible threshold factors that must play an additional role when choosing between alternatives like in Figure 3.5.

First, let us assume that each choice implies a given failure probability as perceived by X (and let's call this: 'hazard' or 'danger'), and a given 'threat' or 'damage': i.e. a negative utility due to both the failure (the cost of a wasted activity and a missed reward) and the possible additional damages.[13]

Second, we assume that X is disposed to accept a maximum hazard (*Hmax*) in its choices, in a given domain and situation. In other words, *there is a 'hazard' threshold over which X is not disposed to pursue that choice*.

We are considering the case of delegation branch ($DoT_{XY\tau}$, $U(X)_{d^-}$, $U(X)_{d^+}$), but the same concepts are valid in the case of X's performance (substituting $DoT_{XX\tau}$, $U(X)_{p^-}$, $U(X)_{p^+}$). In Figure 3.6 we have:

$H_{perceived}$ is the failure hazard perceived by X;

H_{max} is the maximum failure hazard acceptable by X;

σ_H is the hazard threshold.

[13] Thus here we will use the term 'risk' as the result of the entity of losses (damage or threat) and of its probability (hazard or danger). Risk theory (Kaplan and Garrik, 1980) calculates the risk as the product of uncertainty (subjective probability) and damage; other authors propose – for the objective risk – the product of frequency and magnitude of the danger. We are interested in the subjective dimension, so risk should be in our terminology hazard * damage. (Common sense would prefer to call 'risk' the probability, and 'danger' the global result of probability and damage).

To choose a given path it is necessary that:

$DoT_{XY\tau} \geq \sigma_H = (1\text{-} H_{max})$

We claim that such a threshold can vary, not only from one agent to another (*personality*) but also depending on several factors in the same agent. In particular, we claim that the acceptable hazard varies with the importance of the *threat-damage* and with the expected reward. In other words, σ_H (where $0 \leq \sigma_H \leq 1$) is a function of both ($U(X)_{d^-}$) and ($U(X)_{d^+}$): $\sigma_H = f(U(X)_{d^-}, U(X)_{d^+})$.

More precisely: the greater the damage ($U(X)_{d^-}$) the more it grows σ_H; while the greater the utility of the potential achievements ($U(X)_{d^+}$) the more σ_H is reduced.

Moreover, we may also introduce an *'acceptable damage' threshold* σ_d: it fixes the limit of the damage X can endure. Under this value the choice would be regarded as unacceptable.

We have also introduced a *minimal acceptable value* for $U(X)_{d^+}$ (σ_a, *payoff threshold*): under this value the choice would be considered inconvenient.

The function σ_H is such that when $U(X)_{d^-}$ is equal (or lesser) than σ_d then σ_H is equal to 1 (in practice, that choice is impossible).

At the same time we can say that when $U(X)_{d^+}$ is equal (or lesser) than σ_a then σ_H is equal to 1. For each agent both σ_d and σ_a can assume different values.

One might also have one single dimension and threshold for *risk* (by using the formula 'damage * hazard'). However, we claim that there could be different heuristics for coping with risk (this is certainly true for human agents). For us, a great damage with a small probability and a small damage with a high probability do not necessarily represent two equivalent risks. They can lead to different decisions, they can pass or not pass the threshold.

To go back to the case of delegation branch (it is sufficient to substitute $U(X)_{d^-}$ with $U(X)_{p^-}$, $U(X)_{d^+}$ with $U(X)_{p^+}$, to obtain the case of X's performance branch) we have:

$$\sigma_H = f(U(X)_{d^-},\ U(X)_{d^+}) \text{ and in particular}$$
$$\sigma_H = 1 \text{ when } (U(X)_{d^-} \leq \sigma_d) \text{ or } (U(X)_{d^+} \leq \sigma_a).$$

In other words, we assume that *there is a risk threshold* – more precisely a *hazard* threshold depending also on a *damage* threshold – under which the agent refuses a given choice even if the equation (3.4) suggests that choice is the best. It might be that a choice is convenient (and the best) for the ratio between possible payoff, costs and risk, but that the risk *per se* is too high for that agent in that situation.

Let us consider an example (Figure 3.7):

Given $U(X)_{p^+} = 10$, $U(X)_{p^-} = 1$, $U(X)_{d^+} = 50$, $U(X)_{d^-} = 5$, and $Dot_{XY\tau} = Dot_{XX\tau} = 0.7$, the equation (3.5) is satisfied: 0.70 > (0.70 * 9/45) + (4/45) = 0.23. So on the basis of this equation agent X should delegate the task τ to agent Y. However, suppose that the maximum acceptable damage for X is $\sigma_d = 4$ (the damage grows as the $U(X)_{d^-}$ is reduced) then the choice to delegate is stopped from the saturation effect.

Vice versa, considering the example in Figure 3.8, with $Dot_{XY\tau} = 0.7$ and $Dot_{XX\tau} = 0.1$, the equation (3.4) is also satisfied: 0.7 > (0.1*10/11) + (6/11) = 0.63. Also, again on the basis of this equation, agent X should delegate the task τ to agent Y. But if the *minimal acceptable value* for delegation is $\sigma_a = 18$, then the choice to delegate is stopped from the saturation effect because it is considered unconvenient.

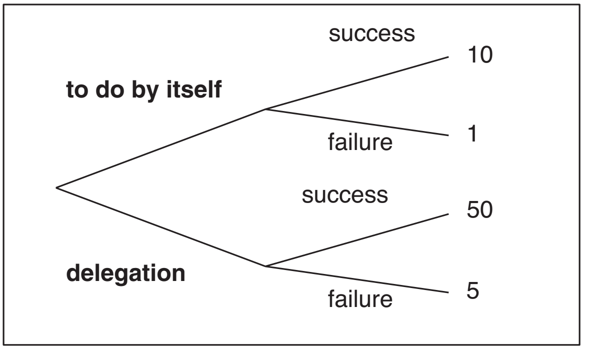

Figure 3.7 An Example in which the maximum acceptable damage is over the threshold. (Reproduced with kind permission of Springer Science+Business Media © 2001)

It is possible that all the branches in the decision scenario would be in a situation of saturation (σ_H=1). What choice does the agent have to make? In these cases there could be several different possibilities.

Let us consider the scenario in Figure 3.5. There could be at least four possibilities:

1) Saturation due to σ_d for branch 'to do by itself' (the potential damage is too high); saturation due to σ_a for branch 'delegation' (the payoff is too low).
2) Saturation due to σ_a for branch 'to do by itself' (the payoff is too low); saturation due to σ_d for branch 'delegation' (the potential damage is too high).
3) Saturation due to σ_a for branch 'to do by itself' (the payoff is too low); saturation due to σ_a for branch 'delegation' (the payoff is too low).
4) Saturation due to σ_d for branch 'to do by itself' (the potential damage is too high); saturation due to σ_d for branch 'delegation' (the potential damage is too high).
   - In the cases (1) and (2) the choice could be the minimum damage (better a minor payoff than a high damage).
   - In the case (3) if $(\sigma_a - U(X)_{d^+}) < (\sigma_a - U(X)_{p^+})$ then the choice will be 'to do by itself' and vice versa in the opposite case. In other words, given a payoff threshold greater than both the utilities (performance and delegation) the action with greater utility (although insufficient) will be better.

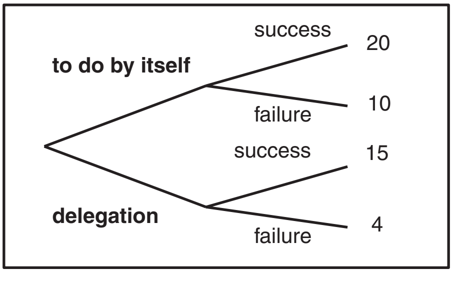

Figure 3.8 An Example in which the *minimal acceptable value* is under the threshold

- In the case (4) if $(U(X)_{d^+} - \sigma_d) > (U(X)_{p^+} - \sigma_d)$ then the choice will be 'to do by itself' and vice versa in the opposite case (the choice is the minor damage).
- In the cases (3) and (4) if $(\sigma_{a^-}\ U(X)_{d^+}) = (\sigma_{a^-}\ U(X)_{p^+})$ and $(U(X)_{d^+} - \sigma_d) = (U(X)_{p^+} - \sigma_d)$ then will the equation (3.5) be the one that will decide what is the right choice.

3.6 Generalizing the Trust Decision to a Set of Agents

It is possible to determine a trust choice starting from each combination of credibility degrees – $\{DoT_{Ag1,\,Agi,\,\tau}\}$ with $Ag_i \in \{Ag_1, \ldots, Ag_n\}$ – of the main beliefs included in Core-Trust and Reliance of Ag_1, and from a set of Ag_1*'s* utilities $\{U_{p^+}, U_{p^-}, U_{di^+}, U_{di^-}, U_0\} = U(Ag_1)$, with $i \in \{2, \ldots, n\}$.

It is possible that – once fixed the set of utilities and the kind and degree of control – different combinations of credibility degrees of the main beliefs produce the same choice. However, in general, *changing the credibility degree of some beliefs should change the final choice about the delegation* (and the same holds for the utilities and for the control).

So, if we suppose we have a set constituted by: $\{DoT_{Ag1,Agi,\tau}\}$ and $U(Ag_1)$, we will have, as a consequence, the delegation $Kind_0Delegates(Ag_1\ \ Ag_i\ \ \tau_0)$ with $Ag_i \in \{Ag_2, \ldots, Ag_n\}$, and $Kind_0Delegates \in$ {*Performs, Weak-Delegates, Mild-Delegates, Strong-Delegates, Nothing*}[14].

At a different time we might have a new set $\{DoT_{Ag1,Agi,\tau}\}$ and/or a new set of utilities.

In Chapter 7 (about the delegation adjustments) we will see how, in order to adjust a given delegation/adoption, it is necessary for the agent to have specific reasons, that is new beliefs and goals. What, in fact, this means, is simply that:

- the delegator's mental state has changed in at least one of its components in such a way that the action to choose is different from the previous one; or
- the delegee's level of self-trust or the delegee's trust in the environment has changed, and there is some disagreement with *the delegator* about this.

At the same time the new sets of beliefs and utilities might suggest various possible strategies of recovery of the trust situation: i.e. given $Kind_0Delegates(Ag_1\ \ Ag_i\ \ \tau)$, $DoT_{Ag1,Agi,\tau}$ and $U(Ag_1)$ we might have an adjustment of $Kind_0Delegates$ (for example from *Weak-Delegates* to *Strong-Delegates*).

This adjustment reflects a modification in the mental ingredients. More precisely, the trustor/delegator either *updates or revises* their delegation beliefs and goals, i.e.:

- either she revises its *core trust beliefs* about *the trustee/delegee* (the latter's goals, capabilities, opportunities, willingness);
- or she revises its *reliance beliefs* about: i) her dependence on *the trustee/delegee*, or ii) her preference to delegate to *trustee/delegee* than to do it herself, or to delegate to Ag_3 (a third agent) or to renounce the goal;
- or she changes her risk *policy* and more or less likely she accepts the estimated risk (this means that the trustor changes either her set of utilities ($U(Ag_1)$)) or her set of thresholds.

[14] For the meaning of *Weak*, *Mild* and *Strong Delegation* see Chapter 2. *Perform* means that the trustor does not delegate but personally performs the task.

In other words, either Ag_1*'s* trust of Ag_2 is the same but her preferences have changed (including her attitude towards risk), or Ag_1 has changed her evaluations and predictions about relying on Ag_2. Another important role is played by the control (see Chapter 7) that can allow delegation also to be given to a not very trusted agent. For an analysis on this relationship see (Castelfranchi and Falcone, 2000)).

3.7 When Trust Is Too Few or Too Much

Trust is not always rational or adaptive and profitable. Let's see when it is rational or irrational, and when it is not useful, although well grounded.

3.7.1 Rational Trust

In our view trust can be rational and can support rational decisions. *Trust as attitude* (*core Trust*) is epistemically rational when it is reason-based. When it is based on well motivated evidence and on good inferences, when its constitutive beliefs are well grounded (their credibility is correctly based on external and internal credible sources); when the evaluation is realistic and the esteem is justified, not mere faith.

The *decision/action of trusting* is rational when it is based on an epistemically rational attitude and on a sufficient degree relative to the perceived risk. If my expectation is well grounded and the degree of trust exceeds the perceived risk, my decision to trust is subjectively rational.[15]

To trust is indeed irrational either when the accepted risk is too high (relative to the degree of trust), or when trust is not based on good evidence, is not well supported. Either the faith component (unwarranted expectations) or the risk acceptance (blind trust) are too high.[16]

3.7.2 Over-Confidence and Over-Diffidence

Trust is not always good – also in cooperation and organization. It can be dangerous both for the individual and for the organization. In fact the consequences of over-confidence (the excess of trust) at the individual level are: reduced control actions; additional risks; non careful and non accurate action; distraction; delay in repair; possible partial or total failure, or additional cost for recovering.

The same is true in collective activity. But, what does 'over-confidence' i.e. excess of trust actually mean? In our model it means that X accepts too much risk or too much ignorance, or is not accurate in her evaluations. Noticed that there cannot be too much positive trust, or esteem of Y. It can be not well grounded and then badly placed: the actual risk is greater than the subjective one. Positive evaluation on Y (trust in Y) can be too much only in the sense that it is more than that reasonably needed for delegating to Y. In this case, X is too prudent and has searched for too much evidence and information. Since also knowledge has costs and

[15] For a more detailed discussion about rational and irrational motive for trust see Chapter 8, Section 8.1.

[16] Rational trust can be based not only on reasons and reasoning, on explicit evaluations and beliefs, but also on simple learning and experience. For example the prediction of the event or result can be based not on some understanding of the process or some model of it, but just on repeated experiences and associations.

utility, in this case the cost of the additional knowledge about *Y* exceeds its utility: *X* already has enough evidence to delegate. Only, in this case, the well-grounded trust in *Y* is 'too much'. But notice that we cannot call it 'over-confidence'.

In sum, there are three cases of 'too much trust':

- More positive trust in *Y* than necessary for delegating. It is not true that 'I trust *Y* too much' but it is the case that I need too much security and information than effectively necessary.
- I have more trust in *Y* than he deserves; part of my evaluations and expectations are faithful and unwarranted; I do not see or do not take into account the actual risk. This is a case of *over-confidence*. This is dangerous and irrational trust.
- My evaluation of *Y* is correct but I'm too risk prone; I accept too much ignorance and uncertainty, or I bet too much on a low probability. This is another case of *over-confidence*, and of dangerous and irrational trust.

Which are the consequences of over-confidence in delegation?

- Delegating to an unreliable or incompetent *Y*.
- Lack of control over *Y* (*Y* does not provide his service, or provides a bad service, etc.).
- Delegation which is too 'open': unchecked misunderstandings, *Y*'s inability to plan or to choose, etc.

Which, on the contrary, are the consequences of insufficient confidence, of an excess of diffidence in delegation?

- We do not delegate and rely on good potential partners; we miss good opportunities; there is a reduction of exchanges and cooperation.
- We search and wait for too many evidences and proofs.
- We make too many controls, losing time and resources and creating interferences and conflicts.
- We specify the task/role too much without exploiting *Y*'s competence, intelligence, or local information; we create too many rules and norms that interfere with a flexible and opportunistic solution.

So, some diffidence, some lack of trust, prudence and the awareness of being ignorant are obviously useful; but, also, trusting is at the same time useful. Which is the right ratio between trust and diffidence? Which is the right degree of trust?

- The right level of positive trust in *Y* (esteem) is when the marginal utility of the additional evidence on *Y* (its contribution for a rational decision) seems inferior to the cost for acquiring it (including time).
- The right degree of trust for delegating (betting) is when the risk that we accept in the case of failure is inferior to the expected subjective utility in the case of success (the equation is more complex since we have also to take into account alternative possible delegations or actions).

3.8 Conclusions

What is to be noticed in this chapter is how one can derive a precise model of the 'degree' of trust simply from independently postulated beliefs, expectations, evaluations, and their properties (like the 'certainty' of the belief, or the quantity of the quality or virtue). In general, in our model of cognition, the pursued goals, the intentions, are based on beliefs, on 'reasons'.

This is also why trust and trust decision can be the object of argumentation and persuasion: I can provide you with reasons for trusting or not trusting *Y*; I can convince you. Of course, trust can also be the result of mere suggestion, of manipulation, of attraction, and other affective maneuvers (see Chapter 5); but here we were modeling explicit and arguable trust.

It is also important to notice that the impact of such a trust degree in decision making is not just due to the 'expected utility'; the process is more complex: there are specific thresholds, there are differences between high probability and low value versus low probability and high value.

It is also important not to have simplistic models of trust degree in terms of mere statistics or reinforcement learning; or of trust decision in terms of delegating to the most trustful guy. An important additional sophistication we should have introduced – at least for modeling human trust – would be the asymmetric evaluation of gains (and missed gains) and of losses (and avoided losses), as explained by 'Prospect Theory': the same amount of money (for example) does not have a comparable impact on our decision when considered as acquisition and when considered as loss; and, as for losses, we are risk prone (we prefer uncertain losses to certain losses), while for winnings we are risk averse (prefer certain winnings to uncertain ones) (Allais, 1953).

In sum, trust (as *attitude and disposition*) is *graded* for seven different reasons:

1. Because it is based on explicit beliefs (like 'evaluations') with their degree of subjective *certainty*, recursively due to trust in evidences and sources: on such a basis, *X* is more or less sure, convinced that, and so on.
2. Because it is based on implicit, felt 'beliefs': sensations, somatic markers, emotional activations, with their *intensity* and affective qualities (safety, worry, etc.); the functional equivalents of 'beliefs' and explicit evaluations.
3. Because those judgments are about *Y*'s *qualities*, virtues, and they can be gradable: *Y* can be more or less skilled, or competent, or persistent, etc. In other words, trust is graded because *trustworthiness* is graded.
4. Because it is *multi-dimensional* (and trustworthiness too); and the global judgment or feeling is the combination of those dimensions.
5. Because it is relative to some Goal of *X*'s, and goals have a '*value*': they are more or less important.
6. Because it is a prediction about a future event, and thus about a subjective *probability* of such an event.
7. Because it presupposes some *risks* (both failure, costs, and possible dangers), that might be perceived with some tangible amount and threshold.

As *decision and act*, trust can be more or less convinced and sure, but cannot really be graded, since *X* has to decide or not, given some threshold of risk acceptance and convenience.

References

Allais, M. (1953) Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine, *Econometrica* 21, 503–546.

Castelfranchi, C. and Falcone R. (2000) Trust and control a dialectic link, *Applied Artificial Intelligence Journal*, Special Issue on Trust in Agents. 14 (8).

Dempster, Arthur P. (1968) *A Generalization of Bayesian Inference*, Journal of the Royal Statistical Society, Series B, Vol. 30, pp. 205–247.

Hunt, D. P., and Hassmen, P. (1997) *What it means to know something*. Reports from the Department of Psychology, Stockholm University, N. 835. Stockholm: University of Stockholm.

Jennings N. R. (1993) Commitments and conventions: The foundation of coordination in multi-agent systems. *The Knowledge Engineering Review*, 3: 223–250.

Kaplan, S. and Garrick, J. (1980) On the quantitative definition of risk. In *Risk Analysis*, 1 (1).

Luhmann, N. (1990) Familiarity, confidence, trust: Problems and alternatives. In D. Gambetta (ed.), *Trust* (Chapter 6, pp. 94–107). Oxford: Basil Blackwell.

Marsh, S.P. (1994) Formalising Trust as a computational concept. PhD thesis, University of Stirling. Available at: http://www.nr.no/abie/papers/TR133.pdf.

Pears, H. (1971) *What is Knowledge ?* New York, Harper and Row.

Sichman, J. R., Conte, C., Castelfranchi, Y., Demazeau. A social reasoning mechanism based on dependence networks. In *Proceedings of the 11th ECAI*, 1994.

Shafer, G. (1976) *A Mathematical Theory of Evidence*, Princeton University Press.

Snijders, C. and Keren, G. (1996) Determinants of Trust, Proceedings of the workshop in honor of Amnon Rapport, University of North Carolina at Chapel Hill, USA, 6-7 August.

von Neumann, John and Morgenstern, Oskar (1944) *Theory of Games and Economic Behavior*, Princeton University Press.

Williamson, O. E. Calculativeness, trust, and economic organization, *Journal of Law and Economics*, XXXVI: 453–486, April, 1993.