# 4

# The Negative Side: Lack of Trust, Implicit Trust, Mistrust, Doubts and Diffidence

In this chapter we analyze the theory of lack of trust, mistrust, diffidence, and the nature of pessimism and optimism in 'giving credit' and entrusting, and also the implicit forms of trust, not based on explicit evaluations and specific beliefs (putting aside trust as feeling and affect – see Chapter 5). These are fundamental issues, not to be theoretically simplified or just reduced to a trivial technical trick.

## 4.1 From Lack of Trust to Diffidence: Not Simply a Matter of Degree

Between full trust and absolute mistrust there is not just a difference of quantity and a continuum from *0* (*complete lack of trust*) to *1* (*full trust*) (see Figure 4.1); like for scholars just considering 'trust' as subjective probability of a favorable event.

Neither is it a value between $-1$ (negative 'trust', the complete mistrust) and $+1$; where *0* would be the simple lack of trust. As already argued by Ulmann-Margalit (Ulmann-Margalit, 2001), however, without a complete and formal systematization of the issue – *trust and its negative counterparts are qualitatively different mental states*.

There exist diverse forms and kinds of lack of trust, which are not just a matter of 'degree' or 'intensity', but must be analyzed in their specific ingredients.[1]

Actually we have to characterize *five* different states (*seven*, if we consider also 'diffidence' and '(not)giving credit') and complicated relations of conceptual or extensional inclusion or exclusion among them.

[1] Also because trust is in part based on the judgments about features and 'qualities', on evaluations. As beliefs – as we saw - they can have a 'degree' (the strength of subjective certainty), but as qualities or 'signals' not all of them are 'gradable': for example, 'PhD' or 'married' or 'nurse'.

*Trust Theory: A Socio-Cognitive and Computational Model* Cristiano Castelfranchi and Rino Falcone
© 2010 John Wiley & Sons Ltd. ISBN: 978-0-470-02875-9

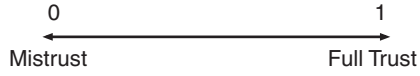


**Figure 4.1** From Mistrust to Full Trust

Let us put aside – for the moment – 'quantities' and the additional problem of 'uncertainty' and of giving or not 'credit' (see below). Let us consider the issue from the point of view of the theory of evaluation (given that – as we have shown – trust basically is a matter of 'evaluation') and of the theory of beliefs (given that explicit trust is grounded on a set of beliefs). In a belief-based approach 'mistrust' is not simply *0* trust or negative *(−1)* trust (Marsh, 2005), but it is the concrete presence of a negative evaluation about *Y*, with its possible supports, and/or of a negative expectation.

### *4.1.1 Mistrust as a Negative Evaluation*

Let's start from the belief that '*Y* is NOT reliable; one cannot trust him':

$$Bel_X\,(Not\,(Trustworthy\;Y)) \tag{4.1}$$

We will call this belief 'mistrust' (and its consequences): a specific judgment about *Y*, a relevant negative evaluation of *Y*, not simply a lack of relevant positive evaluations. If this negative evaluation/expectation is specific (for example: 'If we delegate him without controlling him, everything will be really indecent!') and possibly based on specific 'qualities' (for example: 'He is a dirty and untidy guy'), it is even clearer that we are not simply speaking of degrees of certainty and/or of a value in a continuum.

There are two kinds of negative evaluations (Miceli, 2000) (Castelfranchi, 2000).

- *Inadequacy evaluation*

  '*Y* is *not* able to, is *not* good, apt, useful, adequate, . . . for my goal';

$$Bel_X\,(Not\,(Good\text{-}For\;Y\;g)) \wedge (Goal_X g)\,[2] \tag{4.2}$$

- *Nocivity/dangerosity evaluation*

  '*Y* is 'good' (powerful) but for bad things (contrary to my goal), it is noxious, dangerous'

$$Bel_X\,(Good\text{-}For\;Y\;g')) \wedge (Goal_X\;Not(g')) \tag{4.3}$$

[2] *GoodFor* is the evaluative form/use of the concept of 'Power of' (Chapter 2).

Correspondingly, there are two kinds of 'negative trust' or better of 'mistrust'; two opposites of trust: mistrust in *Y* as for being good/able/reliable/..; suspicion, worry, and diffidence towards *Y*:

- ***Negative Trust-1*** or ***Mistrust*** it is when *X* believes that *Y* is not competent or able, or that his behavior is not predictable and reliable. *X* isn't afraid of *Y* being malevolent, indifferent, irresponsible, or with other defects; she is just convinced that *Y* cannot/will not realize her goal *g* effectively. He doesn't have the needed qualities and virtues (*powers of*).
- ***Negative Trust-2*** or ***Diffidence/Suspicion*** it is more than this. At the explicit level it is some sort of 'paradoxical trust': *X* believes that *Y* has powers and abilities, and also that he can intend and realize something; but something bad (for *X*)! *X* is not simply predicting a failure (*X* cannot rely on *Y*), but probable harms from *Y*, because of *Y*'s bad powers or dispositions.[3]

## 4.2 Lack of Trust[4]

Both the previous forms of negative trust clearly imply 'lack of trust', since the belief that *'Y is NOT reliable for g; one cannot trust him (as for. . .)'* (*formula (4.1)* is true) – which logically excludes that *'Y is reliable'* – is just a sub-case of the mental state: '*X* does NOT believe that *Y* is reliable/trustworthy' (where 'Not ($Bel_X\ q$)' denotes the absence of such Belief in *X*'s mind):

$$Not\,(Bel_X\,(Trustworthy\ Y)) \tag{4.1a}$$

We actually call this mental state 'lack of trust'.

This belief – per se – is clearly compatible with and can also cover another idea: '*X* does NOT believe that *Y* is reliable/trustworthy' *Not* ($Bel_X$ *(Not (Trustworthy Y)))*. This is when *X* doesn't know, has no idea: the case of ignorance or pending judgment. When '*X* does NOT believe that *Y* is trustworthy', either she believes that *Y* is not trustworthy, is unreliable; or she doesn't know how to evaluate *Y*. The 'lack of trust' covers both *mistrust* and *doubt*:

$$Not\,(Bel_X\,(Trustworthy\ Y)) \wedge Not(Bel_X\,(Not\,(Trustworthy\ Y))) \tag{4.1b}$$

What we have just said does not imply that there is trust only when it is enough for entrusting ('delegating' and rely on)[5] or that there is 'lack of trust' or even 'mistrust' when the agent decides not to entrust *Y*. Logical relations should not be mixed up with quantitative criteria, which are necessary for the decision. These logical relations of inclusion or incompatibility define some 'spaces'. Within the space of '*X* trusts *Y*' (with some degree) then *X* has no mistrust or doubts about *Y*; however, this does not entail that *X* trusts *Y* enough to entrust him.

---

[3] Perhaps *Y* might even realize *X*'s goal g; but by exposing *X* to serious risks and dangers.

[4] One might prefer – for this case – the term 'no trust' and to limit the expression 'lack of trust' just for case 4.1, where trust is not 'sufficient'. However, actually this does not correspond to the current use of the expression (we try to be as coherent as possible). It is in fact normal to say something like: *'But this is lack of trust!'* when for example X has delegated Y but is full of doubts and worries, and would like to continuously check Y's work.

[5] It is a typical, correct, non-contradictory expression to say: *'I trust Y, but not enough'*. This is because 'I trust *Y*' can mean two different things: (i) 'I have a certain amount of trust (positive evaluations) about *Y*'; or, (ii) 'I have a sufficient amount of trust in *Y* to entrust him, rely on him'. This thanks to the pragmatic implications of the sentence 'I trust *Y*' and to the meaning of 'to trust' not just as disposition but as decision and action (that presupposes the 'sufficiency' of the evaluation).

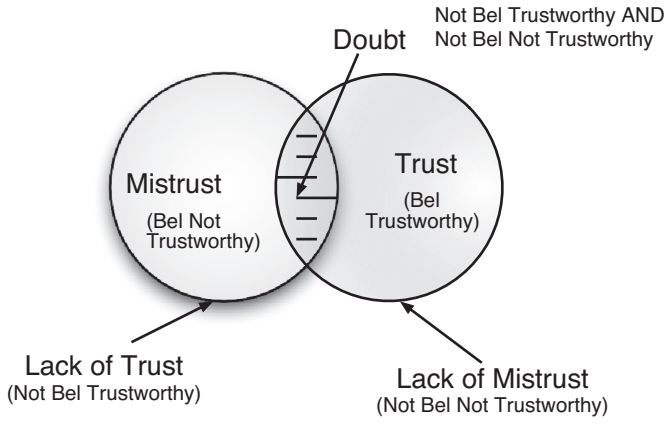


**Figure 4.2** Trust, Mistrust, Lack of Trust, Lack of Mistrust, Suspect, Doubt

Correspondingly, within the 'space' of lack of trust, clearly there is no trust. However, it is perfectly possible and usual that *X* 'in part' trusts *Y* (and perhaps enough to delegate to *Y*) and 'in part' doesn't trust or even mistrusts *Y*.[6]

More than this: *any trust implies some degree of mistrust or at least a lack of trust and doubt*. Since we do not admit that trust is the total certainty and evidence, and we claim that in any case trust entails some (at least *implicit*) risk of failure or some perceived ignorance. Vice versa, any mistrust too is not certainty, and presupposes some (although minimal) doubt in favor. This is a general theory about beliefs with a degree of certainty, and applies to evaluation, to expectations, etc. For example, any hope – by definition – implies some worry; and any worry (implicitly) contains some hope; although the complementary mental state is not necessarily derived (it can remain just implicit) or focused on. (See Chapter 2).

## 4.3 The Complete Picture

We have also to add: *'X believes that Y is reliable'* ('trust'), which is the only state incompatible with 'lack of trust': *formula (4.1)*. However, this case also implies another belief and is a subcase of its extension: '*X* does NOT believe that *Y* is NOT reliable/trustworthy' (*Not* ($Bel_X$ *(Not (Trustworthy Y))*)).

The complete picture (including also 'suspicion') is shown in Figure 4.2.

Here we can see that – obviously – 'trust' and 'mistrust' exclude each other, while 'lack of trust' and 'lack of mistrust' are not fully incompatible; they partially overlap in an area of 'doubt'. 'Diffidence' is not explicitly represented; it would be a specification within the entire area of 'lack of trust' or of 'mistrust', that is, of the doubt or 'suspicion' that there might be

[6] This does not deny that when we use the expression 'I trust' usually we in practice intend 'I trust enough', while when we use the expression 'I do not have trust' we mean 'I do not have enough trust'. Similarly, when we say 'tall' we actually mean 'taller than the norm/average'. But this is not logically necessary; it is just a pragmatic implication. This is why there is no logical contradiction while saying: 'I distrust, I'm diffident, but not enough for not entrusting him'.

some betrayal or danger (the second kind of negative evaluation). It is an active suspicion, that is, to be alerted, to be vigilant for promptly detecting signs of wrong or dangerous acts of *Y*. This is an *epistemic* attitude which implies epistemic actions (Section 2.3.1), at least to give attention to, to monitor.

As we can see it is a complex picture of various well-defined states. As we said, this becomes even more clear and strong if we assume – like in our model – that our trust-evaluations are based on and supported by specific positive or negative evaluations about *Y*'s characteristics (which actually are other pieces of the trust in *Y*, sub-components of the whole trust judgment, see Section 2.2.1): *Y is expert; Y is persistent; Y is a good friend of mine;* etc. This makes it even more evident that trust and its negative counterparts cannot be reduced at all to a unique dimension and to a simple 'quantity'.

Of course, while reintroducing quantities and degrees the borders and transitions from one case/state to the other (in Figure 4.2) become more fuzzy. What actually does it mean by '*X* does not believe. . . .'? Which degree of certainty is needed for not-believing? 70%? In other words, 'do not believe' because there is an evidence of 70% will be different from 'do not believe' based on an evidence of 60% or 50%.

Moreover, in the intermediate state of the *formula (4.1b)* we can quantify the two components: how much I believe that *Y* is reliable (although believing that he is unreliable) versus how much I believe that he is unreliable (although not believing that he is so). There can be different degrees and thus various kinds of 'doubt' more or less favorable to *Y*.

The stronger it is $Not\,(Bel_X\,(Not\,(Trustworthy\ Y)))$ the closer *X* is to 'trust' $(Bel_X\,(Trustworthy\ Y))$. In sum, the discrete *YES/NO* character of the 'decision' to believe and do not believe makes discontinuous the spectrum of our evaluation of trustworthiness.

## 4.4 In Sum

With the conceptual approach we can characterize *four* different basic mental states relative to perceive *Y*'s trustworthiness, but not a continuum:

$$(\textbf{\textit{Mistrust}}) \quad Bel_X\,(Not\,(Trustworthy\ Y)) \tag{4.1}$$

which also includes *diffidence* if not *suspicion* (that mainly is about *Y*'s intentions, motivations, good disposition).

$$(\textbf{\textit{Lack of trust or no-trust}}) \qquad Not\,(Bel_X\,(Trustworthy\ Y)) \tag{4.1a}$$

$$(\textbf{\textit{Lack of mistrust or No Mistrust}}) \qquad Not\,(Bel_X\,(Not\,(Trustworthy\ Y)) \tag{4.1c}$$

$$(\textbf{\textit{Trust}}) \qquad Bel_X\,(Trustworthy\ Y) \tag{4.1d}$$

(Ulman-Margalit, 2001) is right about asymmetries and implications of those different states, but actually they are merely pragmatic; rules and presuppositions of communication and its implications. On the conceptual and logical plan:

(4.1) is incompatible with (4.1d) and (4.1c), while it is compatible with (4.1a): (4.1a) covers (4.1) and (4.1) logically entails (4.1a). In order to characterize state

(4.1a) without (4.1) one should say: (4.1a) + (4.1c) = (4.1b); that is, pending judgment: neither trust nor mistrust.

(4.1a) is incompatible with (4.1d), but compatible with (4.1) and (4.1c);

(4.1c) is incompatible with (4.1) and with (4.1d), but compatible with (4.1a)

(4.1d) is incompatible with (4.1c), that is, (4.1c) can cover (4.1d) too, while (4.1d) logically entails (4.1c) (thanks to the principle of 'non-contradiction').

## 4.5 Trust and Fear

'Fear' is not just the expectation of a danger or harm, neither is it the emotion due to (elicited by) such an expectation: by the simple explicit representation of a possible harm. To feel fear the danger should be 'significant', not just explicitly taken into account: the subjective expected (dis)utility must be greater than the personal and contextual threshold of acceptable 'risk'. A perceived risk that cannot be coped with and is unbearable, elicits fear (Lazarus, 1991). In fact, every moment we consider (view) possible harms or risks in our situations and decisions, but not all of them elicit 'fear' as a specific emotion.

'Fear' is the extreme lack of trust, is a possible part and reason of 'distrust': *X* not only believes that *Y* is unwilling or unable to do as needed, and one cannot rely on him; but, in particular, *X* believes that *Y* is untrustworthy because he is dangerous: either out lack of attention or through a bad disposition, character, or out of hostility, envy, etc. *Y* can produce some harm to *X* (even intentionally), especially if *X* is exposed and unprotected because of her trusting attitude; some harm going beyond the simple unattainment of the delegated task (failure).

So 'fear' usually is one basis for 'distrust' and is difficult (but not impossible) when it coexists with a decision to trust. Sometimes *X* – although perceiving her exposition and vulnerability, and considering a risk and even feeling some fear – either has no alternative for $\tau$ or believes *Y* to have the willingness and competence to do $\tau$, that she decide to trust *Y* and rely on him, in spite of having some worry and fear. As we said, trust is a bet, necessarily entails some risk, and frequently implies some perceived risk.

## 4.6 Implicit and by Default Forms of Trust

Given the two different opposites of trust (the two kinds of negative evaluations), one can identify an important implicit form of trust where *X* neither has specific and explicit positive evaluations about *Y*, nor has she explicit suspects and worries. She is just *without doubts, suspect and worries*, she naively relies upon *Y*: not because of an explicit and reason based evaluation of *Y*. She trusts by default and because she has no reason to be cautious and to suspect, does so without any explicit examination of whether *Y* is able, willing, or dangerous. One could represent this attitude as the absence of mistrust, of suspicion, but also of explicit positive evaluation and trust.

This implicit, passive, and spontaneous or naive form of trust consists of not having the typical trust beliefs, but also in *not having* negative ones, negative expectations: to be without any alarm and suspicion. Consider in fact that not having a given belief (*I do not believe that*

*it is raining*) is a very different cognitive state than having the opposite belief (*I believe that it is not raining*).

As not trusting *Y* for *g* is not the same as expecting harm from *Y*, analogously not dis/mistrusting, not worrying, is not the same as positively believing that *Y* is capable and willing. However, this lack of dis/mistrust can be sufficient for relying upon *Y*. It depends on the agent's rule. If the agent has a default which in order to delegate requires specific positive evaluations, specific trustworthiness beliefs, the absence of these beliefs actually is distrust. On the contrary, if the agent's default rule is '*except you have specific reasons for not relying on Y, specific negative evaluations, then trust Y*', the lack of mistrust is factually an (implicit) form of trust.

In this weak form the behavior/action of trust consists of the absence of cautions, of controls, of any search for evidence for evaluation, and in the absence of a true 'decision' about trusting or not *Y*. Only after some negative unexpected experience, this kind of trust is damaged. Whoever uses explicit, evaluation-based trust, based on evidence, is no longer naive: she has already considered the situation as problematic; she has some doubt. There is, on the contrary, a form of trust without and before any question like: 'Can/should I trust *Y*?' (See also Chapter 5).

It is important also to distinguish between *uncertainty* (the fact that we do not have complete evidence of our positive evaluation of (*trust in*) *Y*, we are not 100% sure of our beliefs), that make *Y*'s behavior (and results) not completely subjectively predictable; from the actual presence of contrasting, negative evaluations and expectations. The absence of a belief is a mental state significantly different from the presence of the negative belief, with completely different consequences at the reasoning and at the pragmatic level. When *X* has positive evaluations of *Y*, and *does not have any negative (pertinent) evaluation,* although this positive evaluation leaves some room for ignorance and uncertainty, this is very different from a situation where *X* has negative beliefs about *Y* which make *Y* 'ambivalent' (attractive and repulsive, positive and negative, at the same time) and destroys *X*'s 'trust in' *Y*, his trustworthiness. Non-ambivalent although uncertain evaluation is very different from ambivalent evaluation. Thus, we have to distinguish between two types of 'unharmfulness': 'safety' and 'there is nothing to worry' etc.: the implicit and the explicit.

*Implicit un-harmfulness* simply consists of the absence of suspicions, doubts, reasons to worry, diffidence, no perceived threats; some sort of 'by default' naive and non-arguable confidence. I do not have reasons to doubt *Y*'s pro-attitude (active or passive adoption), I do not have negative beliefs about this.

*Explicit un-harmfulness* consists of explicit beliefs about the fact that 'I have nothing to worry from *Y*'.

Both, the implicit or explicit un-harmfulness can be based on other beliefs about *Y*, like *'He is a friend of mine', 'I'm likeable', 'I feel his positive emotional disposition'* (empathy), '*He is honest and respectful of norms and promises', 'He fears me enough'*, . . . and also *'He trusts me and relies on me'*.[7]

Another important kind of 'implicit trust' is the procedural, automatic trust, or better 'confidence', based on perceived regularities, learning, and confirmation of practices.

For example, in motor behavior there are a lot of implicit 'expectations' about objects, movements, etc. and their 'reliability'. And when we (as expected) successfully perform a

[7] This unharmfulness perception and then trust in *Y* based on *Y*'s trust in *X*, is important for the circular dynamics of trust and to explain how trust can create trust (Chapter 6).

given action we unconsciously confirm a lot of implicit assumptions and predictions. We are 'confident' in the ground, in a chair, in a key, etc. Only when something goes wrong, we become suspicious and worry about something. But this form of trust is fundamental also in social life and for the maintenance of social order, which is based on trust and is the basis for trust (Chapter 9).

In sum, the lack of *explicit* trust covers three different mental states:

- *insufficient trust* (*X* does not estimate enough *Y* to count on him, she has some negative evaluation on *Y*) (see below);
- *mistrust* (*X* worries about *Y*);
- *implicit trust*, be it either spontaneous, naive and by default (lack of suspect), or be it automatic and procedural, just based on previous positive experience and learning.

### *4.6.1 Social by-Default Trust*

*By-default trust* is very relevant in social life, especially in communication: we – for example – ask for information from people who we have never met before and we will never meet again, which might deceive us without any external sanction. In general our claim is that Grice's principles about linguistic communication are in fact two default rules. It is obviously true that people can lie, but they should have some reason for doing so; it is true that people can not believe, reject information, but they have to have some reason for this. The speaker's and the hearer's default rules in linguistic communication are as follows; and the hearer rule is a 'trust' rule.

*The speaker's default rule:*

'except you have specific reasons for deceiving, say the useful truth'

*The hearer's default rule:*

'except you have specific reasons for being diffident, believe the speaker'

We justify the existence of this default-rule (to say the relevant truth and to ask for information and believe in people) with a form of 'reciprocal altruism' in humans about exchanging and circulating knowledge (Conte, 1995). Knowledge is such a crucial resource for human beings and is made more valid by social circulation and sharing, that passing on in a reciprocate fashion becomes a serious advantage.

An important form of by-default trust is also in 'generalized trust' (see Chapter 6):

'Except you have specific reasons for diffidence towards a given agent, trust everybody in this context/community/category'

The reason why we are so sensitive to a trust atmosphere is that this is the 'signal' of a diffuse default-rule, and we can adopt it while greatly reducing the cost of search and monitoring of information, and our stress and worries.

In a sense 'by-default' trust is another form of 'conditional' trust. It is a belief of trustworthiness relative to another belief:

$$IF\&\ UNTIL\,(Not\,(Bel_X\,(Not\,(Trustworthy\ Y)) \rightarrow Bel_X\,(Trustworthy\ Y) \tag{4.4}$$

where $\rightarrow$ means 'implies'.

Or it is some sort of conditional rule about the trust act:

$$IF\&\ UNTIL\,(Not(Bel_X\,(Not\,(Trustworthy\ Y)) \Rightarrow Trust\,(X\ Y) \tag{4.5}$$

where $\Rightarrow$ means 'produces'.

This means that $X$ will control the validity of her assumption; and its confirmation.

Of course, this default rule is very different (in feeling and behavior) from the possible opposite one; some sort of *conditional distrust*:

$$IF\&\ UNTIL\,(Bel_X\,(Trustworthy\ Y)) \rightarrow Not\,(Trust(X\ Y)) \tag{4.6}$$

## 4.7 Insufficient Trust

*Trust insufficiency* is when the trust that $X$ nevertheless has in $Y$ does not exceed the threshold (adopted by $X$ in that context for that task) necessary for the decision to delegate, entrust $Y$.

Notice that:

- The amount of trust can be even greater than the amount of distrust or of lack of trust; but, nevertheless, insufficient (see Figure 4.3).
- when trust is not sufficient, lack of trust is too much, but it is not necessarily all 'distrust': there can be a lot of ignorance or doubt (see below).

While introducing more sophisticated and quantitative models, we get other notions that are very interesting, like 'evidence/reason-based' trust/distrust *versus* trust/distrust just based on giving/not-giving credit.

There is another form of trust 'insufficiency': a relative not an absolute one.

Negative evaluations can in fact be 'absolute' or 'comparative' and 'relative' to a standard, threshold, etc. (Chapter 2 and 3). If the evaluation of $Y$ is 'inferior' to the needed threshold or standard, or inferior to the evaluation of $Z$, it becomes 'negative' ($Y$ is insufficient, is inferior). In a sense here we are at a meta-evaluative level; there is a sort of meta-evaluation: an evaluation of the evaluation: '*Is this evaluation sufficient?*' '*No; it is not a really good evaluation*'. In particular, it is not really a 'lack of trust' a trust which is or would be 'enough'

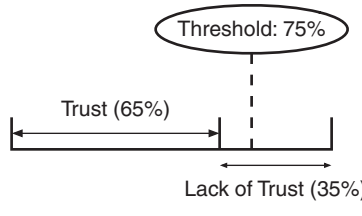


**Figure 4.3** For the Delegation, Trust must be over the threshold

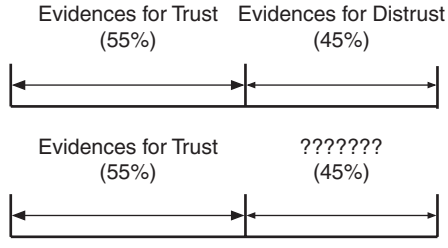


**Figure 4.4** Evidence-based Trust matched with. a) Evidence-based Distrust and b) lack of knowledge

for trusting, but being inferior to another Agent ($Z$) is not winning. $X$ trusts $Y$, she finds $Y$ even reliable and might entrust him, but – since there is also $Z$, who is even better – she will chose $Z$ and (en)trust him. This is an 'insufficient' evaluation of $Y$ but not really a 'lack of trust' in $Y$. Or better, it is a 'relative lack of trust' but not an 'absolute lack of trust'.

## 4.8 Trust on Credit: The Game of Ignorance

Trust (usually) goes *beyond* the evidences and certainty of well-grounded beliefs; in this it consists of 'giving credit' (or not), that is, to believe even beyond the evidence.

Let us adopt a model *à la* Dempster and Shafer,[8] where, between the evidence in favor of a given eventuality that $P$ (thus of the well established probability of $P$) (say 55%) and the evidence in favor of *Not-P* (thus its estimated probability) (say 15%), there is a gap of 'ignorance' (say 30%), a lack of proof in favor or against $P$. First of all, let us observe that with a given level of 'evidence-based' trust (say 55%) it is very different whether the rest, the complement, is supported distrust or just lack of knowledge, just possibility, not probability of *Not-P* (see Figure 4.4).

These two scenarios are psychologically very different: they induce opposite behaviors. For example, in the second scenario the subject might suspend her judgment and decision while waiting or searching for new data; she might perceive a strong uncertainty and be anxious. While in the first situation she might feel able to take a decision (positive or negative).

This is additional confirmation of the fact that it is not sufficient to have just an index, a number (say 55%, or 0.5) to represent trust.

It is important to stress that *with just one and the same level/degree of trust there may be completely different feelings, decisions and behaviors*. Moreover, an 'insufficient' trust can be due to quite different cases: factual distrust, negative evidence, or too much uncertainty and ignorance.

### *4.8.1 Control and Uncertainty*

As we just said, lack-of-trust because of insufficient information is different from lack-of-trust due to supported low evaluations. In the former case, I can wait or search for additional information. Sometimes, it is possible to adopt such a strategy as 'run time', 'work in progress', and then start to delegate (trust is enough!) with the reserve of *monitoring*.

[8] See for example: http://en.wikipedia.org/wiki/Dempster–Shafer_theory

Control is useful and can be a remedy in both cases, but with different roles. Control means in fact (see Chapter 7 for a more detailed analysis):

a) The possibility of improving *Y*'s performance, or of intervening in time – on the basis of new evidence – for remedying, revoking the delegation, etc. Control is a remedy to low expectations about *Y*.
b) The possibility of acquiring, through *monitoring,* additional evidence and information run-time.

In sum, control (monitoring + intervention) is also a remedy against uncertainty-based lack-of-trust: '*In case something is bad I will know, and I will be able to do something*'. I use monitoring in order to *update* and complete my beliefs, since I know that my prediction of the future and my knowledge is incomplete. This is different from a low probability-based lack-of-trust: '*In case some prediction is wrong I will know, and I will be able to do something*'. I use monitoring in order to *revise* my beliefs that were wrong.

Of course, the two problems and the two functions of monitoring can co-occur.

### *4.8.2 Conditional Trust*

Searching or waiting for additional and determinant evidence also means having to formulate a form of 'conditional' trust. *X* trusts *Y* but if and only if/after *Y* has performed a given action, provided a given assurance, or proof. *'Only if he swears on his sons'; 'Only if I can check', 'Only if has this documented experience',* and so on.

$$IF\,(Bel_X\,(Predicate\ Y)) \rightarrow\ (Trust(X\ Y)) \tag{4.7}$$

Where, *Predicate* represents either an act or a feature of *Y*, or a *sign* of such a feature.

The difference between this 'conditional' trust and normal 'evidence-based' trust is just that *X* is waiting for such evidence; but in a sense she has already decided to trust *Y*, provided that the expected evidence will be true.

### *4.8.3 To Give or Not to Give Credit*

The second remark is that it is extremely important for there to be the possibility to mentally ascribe that gap of ignorance (that part without evidence; what is 'possible' but not grounded) in favor or against *P* or *Not-P*.

In fact, given this model, the part that Dempster and Shafer call 'plausibility', the empty part, is in fact ascribable to *P* or to *Not-P*. *P* is probable 55% but *possible* and plausible (not against evidence) up to the 85%! While *Not-P* is probable 15% but plausible up to the 45%. In our opinion, applied to the prediction of the behavior of *Y X* counts on, this gap represents what we call to 'give credit'. If *X* trusts *Y* beyond supported evidences and grounded probabilities, and gives in favor of *Y* all the 'plausibility' space (till the limit of the opposite evidences), then she is 'giving credit' to *Y*; she is trusting (believing) beyond evidence (see Figure 4.5 and Section 4.2.1).

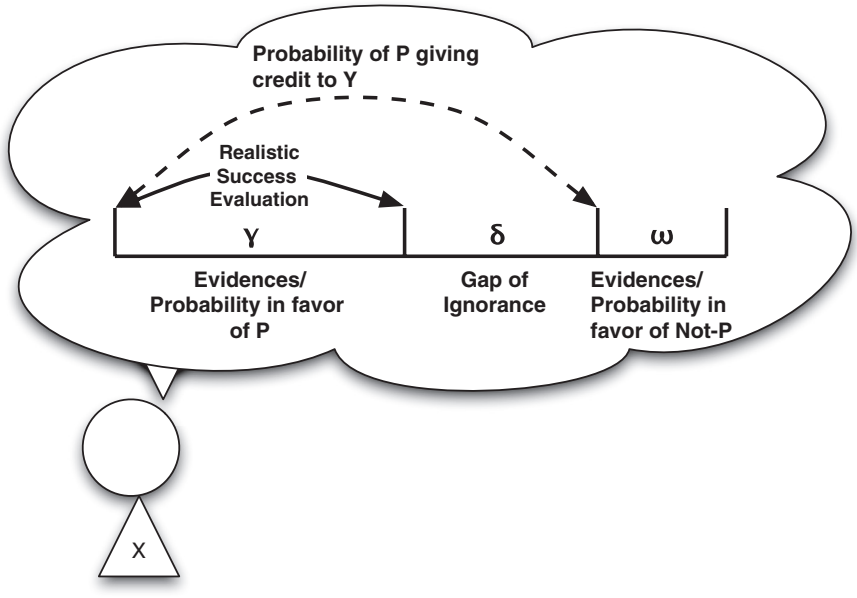


**Figure 4.5** Giving credit

If, on the contrary, *X* has a pessimistic or prudential attitude and decides to trust somebody only after due consideration, only on the basis of evidence and certain data, then the entire space $\beta$ of 'possibility' is assumed in favor of Not-P, as lack of trust (in this frame we will call that space *b*, see Figure 4.5).

*Some scholars would like to call 'trust' only this 'credit', only reliance not based on evidences and good reasons*; but, this is a prescriptive attitude, in contrast with the meaning of the term in various languages, and with the concrete social use of this notion. Also because even proofs, certainty, reasons, are just a subjective fact; a matter of believing, and of trusting sources (there is no way out from such a recursion of trust).

The truth is that *there exist two faces and components of trust* (and of lack of trust): the one '*after due considerations*', that is, given in return for received evidences; the other, given '*on credit*', for free, in view of future confirmations. In both cases, *X* is exposed to risks and has not full certainty; in fact, the 'bet' present in any act of trust is not in trust *per se*, but in counting on the other, in entrusting, in accepting the possibility of a failure or harm.

To 'give trust' in the sense of 'giving credit' is – as stressed by many authors – an *optimistic* attitude[9] not fully prudent and reasonable, but frequently quite effective, and some sort of self-fulfilling prophecy (see Chapter 8).

The inclusion of the theory of ignorance and uncertainty within the theory of trust – which is an additional confirmation that the simple 'subjective probability' of the favorable event is really unsuitable – has interesting consequences.

[9] This is not necessarily due to a personality or stable disposition. It can be due to circumstances, to *Y* and his 'cues' that inspire trust; or to a given context or transitory mood.

In sum, there are three cases, three 'games' played by our ignorance: (i) to play in favor of trust; (ii) to remain indifferent, in between; (iii) to play against trust. (ii) and (iii) are both 'lack of trust', but with two distinguishable faces: unsolved doubts *versus* suspect and diffidence.

On the one hand, *X's* trust in *Y* acquires two 'components': the *positive evaluation part* ($\gamma$), and the *simply credited part* ($\delta$). And there can be quite different proportions between them, for example:

- $\gamma = 0$, that is, pure '*faith*' in *Y*, without any evidence (see note 11 in Chapter 3);
- $\delta = 0$, that is, a stingy trust (perhaps sufficient), mere esteem based on evidences and 'after due considerations', without any additional credit.

It is important to note that $\delta$ trust can be necessary (in some decisions) in order to pass the decision threshold; but probably with a greater perception of *uncertainty*. In our model, in fact, the subjective/perceived 'uncertainty' is the function of two independent dimensions: the amount of ignorance, of lack of data; the balance between the pros and the cons. The greater the lack of data, the *perceived ignorance*, the greater the perceived 'uncertainty'; but also: the smaller the difference between the data in favor of *P* and those in favor of *Not-P*, the greater the perceived 'uncertainty'.

For example, given the same gap of ignorance ('plausibility'), say 60%, our perceived uncertainty is greater if the probability of *P* is 20% and the probability of *Not-P* is 20%, rather than if the probability of *P* is 35% while *Not-P* is just 5%.

These two kinds of uncertainty have a different nature: the first (*due to lack of data*) is about the evidence; the second (*equilibrium between evidence in favor of P and Not-P*) is about the decision to take. However, there is a relationship between them: for example, the first implies the second one.

Our previous claim was simply that the feeling of trust (given a decision and delegation, that is, when $\gamma+\delta$ are higher than the required threshold) will be quite different if $\gamma$ is very consistent or even sufficient and $\delta$ is just additional and reassuring, compared with the situation where $\alpha$ is weak or insufficient and $\beta$ is necessary for the decision. This also means that the *risk perceived* (taken into account) and accepted in the moment of the decision to trust, has two faces: on the one hand, it is the part representing true distrust, *bad evaluations and prediction* ($\omega$); on the other hand, it can just be the *lack of positive evidence*, perceived ignorance, but perhaps given as 'plausibly' in favor.

Not only it is a matter of how much risk *X* is disposed to take (subjectively), but also how blindly *X* accept it.

### *4.8.4 Distrust as Not Giving Credit*

If there is an $\gamma$ (evidence-based) trust and a $\delta$ trust ('there is no evidence but I give it in favor of *P*'), then there also exists a *distrust* of the first kind ($\omega$) and another form of distrust: a 'lack of trust' assumed in a pessimistic attitude as against *P* ($\lambda$).[10] As we said, the 'lack of trust' can cover both 'distrust' in the strict sense (negative evaluations and expectations), and lack

[10] Notice that this operation of 'taking as favorable/good' or 'taking as unfavorable/bad' more precisely is an operation of assumption ('acceptance') than of 'belief'.

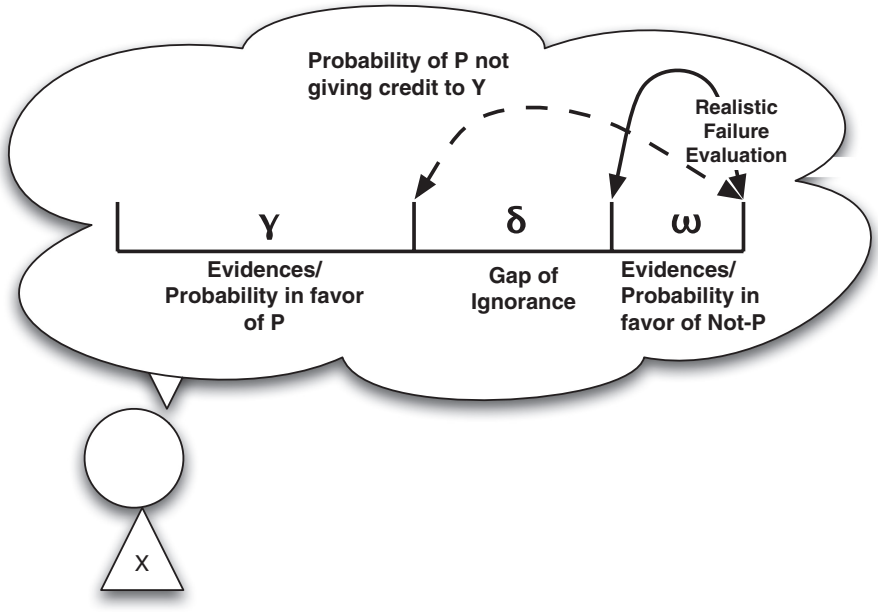


**Figure 4.6** Be prudent

of beliefs and evidence (pros or cons); but also this lack of evidence, assumed as unfavorable, as suspect.

A 'diffident' attitude consists, on the one hand, of being vigilant and alarmed (worrying about possible harms), on the other hand, precisely in being prudent, *not giving credit*, not believing without evidence, being pessimistic and taking as potentially negative all the unclear possibilities (see Figure 4.6).

As we said, the perceived ignorance must not necessarily be ascribed in favor of *P* or *Not-P*. It can be just neutral, as perceived lack of information; and perhaps the decision will be suspended. This entails – as we said – that when trust ($\gamma$) is not sufficient for entrusting, its complement, is mistrust; it can just be perceived ignorance (Figure 4.7).

Notice that in a sense, $\omega$ represents the estimated probability of a failure (of *Not-P*) and thus a crucial component of the perceived 'risk' that might have a specific threshold (Chapter 3). *X* might not accept a given level of $\omega$, and thus renounces on betting on *Y* if the bad evidence on *Y* is too great. Clearly, this is independent of the positive trust and on its acceptance threshold.

For example, if $\omega$ is at the 15% – like in our previous example – it might be acceptable; while if it was at 30% (even still being $\gamma$ at the 55%) the risk might be unacceptable. As we said, *X* may decide not to trust *Y* because trust is not enough, not because distrust is greater than trust, or because all the rest is distrust.

We will develop in Chapter 8 a more complete analysis of 'Optimism' and also the fact that both trust and distrust are actually 'self-fulfilling prophecies', with important consequences on 'rationality' of trust bet, on trust dynamics (Chapter 6), on trust spreading and self-organization.

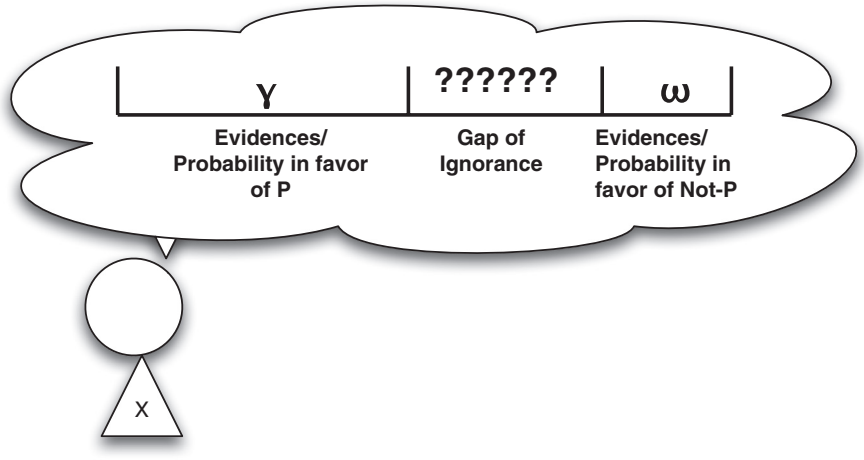


**Figure 4.7** Perceived ignorance and doubt

## References

Castelfranchi, C. (2000) Affective appraisal vs cognitive evaluation in social emotions and interactions. In A. Paiva (ed.) *Affective Interactions. Towards a new generation of computer interfaces* (pp. 76–106). Heidelberg: Springer.

Conte, R. and Castelfranchi, C. (1995) *Cognitive and Social Action*. UCL Press.

Lazarus, R. S. (1991) *Emotion and Adaptation*. New York: Oxford University Press.

Lorini, E. and Castelfranchi, C. (2004) The role of epistemic actions in expectations. In Proceedings of the Second Workshop of Anticipatory Behavior in Adaptive Learning Systems (ABIALS 2004), Los Angeles, 17 July 2004.

Marsh, S., Dibben, M. R. Trust, untrust, distrust and mistrust - an exploration of the dark(er) side, Proceedings of Third *iTrust* International Conference, 2005.

Miceli, M., and Castelfranchi, C. (2000) The role of evaluation in cognition and social interaction. In K. Dautenhahn (ed.), *Human Cognition and Agent Technology* (pp. 225–261). Amsterdam: Benjamins.

Edna Ullmann-Margalit TRUST, DISTRUST, AND IN BETWEEN. Overview1 © Draft September 2001. 'http://www.law.nyu.edu/clppt/program2001/readings/ullman_margalit/Trust,%20distrust%20and%20in%20between.pd'