5

The Affective and Intuitive Forms of Trust: The Confidence We *Inspire*

In this chapter[1] we analyze an aspect of trust which is in a sense 'marginal' relative to our systematic and analytic theory,[2] but not marginal at all from a complete and adequate account of trust and of its real significance and functioning in human interaction: the *affective version or components of trust*. This aspect is also very crucial for social, economic, and moral theories. Is it not actually too 'cold' and reductive to treat trust as a judgment and a reason-based decision? Is not trust something that we just feel and cannot explain?

We have just analyzed the cognitive *explicit* facet of trust as beliefs and goals about something, and a consequent decision of relying upon it. We have completely put aside the affective side: the trust that we 'inspire', the merely intuitive, emotional facet. *It is true that trust can also be this or just this: no judgment, no reasons, but simply attraction and sympathy*. This is an automatic, associative, unconscious form of appraisal: *we do not know why* we prefer *Y* and are attracted by *Y*.

There are beautiful experiments by Bargh's group on this form of affective appraisal. One should also account for the personality aspects of trust as disposition or as default attitude. Some emotions are based on and elicited by true evaluations (beliefs), and also, trust as affective disposition can be based on trust as esteem and good expectations. And the affective aspect of trust can play a role by modifying the belief process, source, and 'decision'. But, on the other hand, trust can be a non-belief-based emotional reaction, an affective attitude simply activated by unconscious perception of signs or associations, by 'somatic markers' (Damasio, 1994), (Castelfranchi, 1998).

[1] This chapter does build heavily on the work with Maria Miceli on the theory of emotions, evaluation, intuitive appraisal.

[2] It is also marginal in part for applications in some domains like ICT (but not 'affective interaction' or web communities), but not at all in other domains like marketing or politics.

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

5.1 Two Forms of 'Evaluation'

As we said trust is a form of appraisal (of the trustee by the trustor); it is an *attitude* based on or implying an evaluation (and an *act* signaling it). However, there are two different forms of evaluation/appraisal in cognitive agents: *explicit evaluative judgments (beliefs)* versus *implicit, affective appraisal*. The first kind has been systematically explained in Chapter 2 (with the theory of 'qualities', 'standards', two kinds of 'negative evaluations' and so on). The other kind deserves some more attention (Miceli, 2000), (Castelfranchi, 2000).

5.2 The Dual Nature of Valence: Cognitive Evaluations Versus Intuitive Appraisal

There are at least two kinds of appreciation of the valence of events, situations, and entities; two kinds of 'evaluation' in a broad sense.

a) A *declarative or explicit form of evaluation*, that contains a judgment of a means-end link, frequently supported by some reason for this judgment, relative to some 'quality' or standard satisfaction.
   This is a *reason-based evaluation* that can be discussed, explained, argued upon. Also the goal of having/using the well-evaluated entity (which is the declarative equivalent of 'attraction') can be 'justified'. This is the classical approach to values (moral or ethical) that is synthesized by the 'motto' (of Aristotelian spirit):

   'it is pleasant/ we like it, because it is good/beautiful'

b) A *non-'rational' (or better non-'reasons-based') but adaptive evaluation*, not based on justifiable arguments; a mere 'appraisal',[3] which is just based on associative learning and memory.

In our view, in the psychological literature on emotions, in particular in the very important and rich literature on emotions as based on a cognitive appraisal of the situation (Frijda, 1986), (Frijda, 1988), (Arnold, 1960), (Scherer, 1986), (Scherer, 1999), there is a systematic and dangerous confusion between these two kinds of 'evaluation' (also in Damasio). Incoherent terms and properties are attributed indifferently to the term 'appraisal' or 'evaluation'. This fundamental forerunner and component of the emotion is characterized – at the same time – as 'cognitive', 'intuitive', 'immediate', 'unconscious', implying also inferences and predictions, etc. We propose (see also (Miceli, 2000), (Castelfranchi, 2000) (Castelfranchi, 2009)) *to distinguish between 'appraisal' - that should be the unconscious or automatic, implicit, intuitive orientation towards what is good an what is bad for the organism- and 'evaluation'.* We reserve this last term (evaluation) for the cognitive judgments relative to what is good or bad for the goal (and why).

[3] Although the English term 'appraisal' is basically a synonym of 'evaluation', let's use it – for the sake of simplicity – for characterizing the second form of evaluation: the intuitive, implicit, affective, somatic, ... appraisal.

5.3 Evaluations

Let us assume there is a good understanding of what 'evaluations' are (Chapter 2, in particular Section 2.2.7), and look in more detail now at the relationships between evaluations, goals, and emotions.

Evaluations imply goals by definition, in that the latter are a necessary component of evaluations, namely, the second argument of the GOOD-FOR predicate. From a more 'substantialist' perspective, evaluations imply goals in the sense that they *originate* from them: it is the existence of some goal *g* (either *X*'s or someone else's) that makes the words good or bad, justifies and motivates both the search for a means *m* to achieve it, and the belief that *m* is (not) GOOD-FOR *g*. Goals and evaluations endow objects and people with 'qualities' and 'faults'.

The relationship between evaluations and goals is even closer, because *evaluations not only implies goals, but can also generate them*. In fact, if *X* believes *m* is good for some goal, and *X* has that goal, *X* is also likely to want (possess, use) *m*. So there is a rule of 'goal generation' which might be expressed as follows: if *X* believes something *m* to be a means for *X*'s goal *g*, *X* comes to have the goal of exploiting/using the means *m*.

Evaluations, that is, knowledge about 'what is good for what', and 'why', play a crucial role in all the cognitive activities that are based upon symbolic and explicit representations, reasoning and deliberation. For example, in problem solving and decision making, the particular advantage offered by evaluative knowledge is precisely a preliminary relationship established between descriptive knowledge and goals, in terms of beliefs about 'what is good for what', derived from either one's experience about problems solved in the past, or one's reasoning and inferences (think for instance of evaluation by standards), or others' communication.

Evaluations make such a relationship explicit; *they fill the gap between knowledge and goals, by 'reinterpreting' the properties, qualities, and characteristics of objects and situations in terms of means for the system's (potential or actual) goals*.

The cognitive network ceases to be neutral and becomes 'polarized' toward goals, that is ready for problem solving and decision-making.

In a cognitive agent preferences can be internally represented both at the *procedural* and at the *declarative* (propositional) level.

- Having a *procedural preference* means that, at a given level of their processing, a system's goals present different degrees or indexes of activation, priority, weight, value, importance (or whatever), that in fact create some rank order among them, which will be followed by some choice/selection procedure.
- Having a *declarative preference* means that the system is endowed with an explicit belief such as: '*m* is better than *n* (for goal *g*)'. In particular, three types of beliefs are relevant for preferences: (a) simple evaluations, that is beliefs about how good/useful/apt/powerful are certain entities relative to a given goal ('*m* is very useful for *g*'; '*n* is quite insufficient for *g*'); (b) comparative evaluations like '*m* is better than *n* for *g*'; (c) reflective preference statements, of the kind 'I prefer *m* to *n* (for *g*)'. Generally, *(b)* are based on *(a)*; while *(c)* are based on *(b)*.

Both procedural and declarative preferences can coexist in a human mind (and would be of some use in artificial minds too), and each level of preference representation – though having its own mechanisms of reasoning – *is translatable into the other*. One can derive a 'weight'

from the declarative evaluations and their arguments, and vice versa, one can explicitly express (as beliefs) some priority of attractiveness, urgency, activation, or whatever.

However, being able to deliberate, that is, to choose an alternative on the grounds of explicit evaluations concerning the 'goodness' of the various options, and being capable of reasoning aimed at supporting such judgments will add further advantages to the mere fact of making choices. In these cases, in fact, the system can *justify* its choices, as well as *modify* the 'values' at stake through reasoning. Moreover, it is liable to persuasion, that is, it can modify its preferences on the grounds of the evaluations conveyed by others (argumentation).

We interact with people on the basis of the image and trust we have of them, i.e. on the basis of our evaluations of them: this defines their 'value' and reputation. And also, social hierarchies are just the resultant of the evaluations that the individuals and the groups receive from others.

5.3.1 Evaluations and Emotions

Given this 'cold' view of evaluation ('cold' if compared with others', e.g., (Mandler, 1975)), what is the relationship between evaluation and emotion? As we claim in (Castelfranchi, 2009):

- *Evaluations do not necessarily imply emotions*

No doubt many evaluations show some emotional feature. For instance, if I believe a certain food, book, person, and so on, to be 'good', I will be likely to feel attracted to it (or him or her). But evaluations and emotions are not necessarily associated with each other, because not any belief about the goodness or badness of something necessarily implies or induces an emotion or an attraction/rejection with regard to that 'something'. There also exist 'cold' evaluations: if, for instance, I believe that John is a good typist, I will not necessarily feel attracted to him. This is especially true because *X* (for example a neutral consultant or expert) can formulate evaluations relative to *Y*'s goals: what would be good or bad for *Y*.

Evaluations luckily have emotional consequences if they simultaneously:

i) are about our own goals (the evaluator is the goal owner);
ii) these goals are currently active;
iii) they are important goals.

- *Emotions do not necessarily imply evaluations*

One may view attraction or rejection for some *m* as a (possible) consequence of an evaluation; so, in this case the emotion 'implies' an evaluation in the sense we have just considered. On the other hand, however, one may view attraction or rejection *per se* as forms of evaluation of the 'attractive' or 'repulsive' object. In the latter case, we are dealing with a supposed identification: to say that an emotion implies an evaluation means to claim that the two actually coincide, which is still to be proved.

In fact, we view attraction and rejection as pre-cognitive implicit evaluation, that we call 'appraisal'. In a sense, any emotion implies and signals an 'appraisal' of its object.

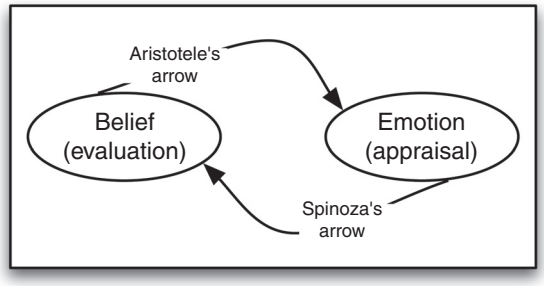

Figure 5.1 From Evaluation to Appraisal and vice-versa

5.4 Appraisal

We assume that a positive or negative emotional response can be associated with some stimulus. The automatic activation of this associated internal response (in Damasio's terms, a 'somatic marker'; (Damasio, 1994)) is the 'appraisal' of the stimulus postulated by several theories of emotions (Arnold, 1960), (Frijda, 1986), (Frijda, 1988), (Lazarus *et al.*, 1970) (Ortony, 1990). The associated negative or positive affective response makes the situation bad or good, unpleasant or pleasant, and we dislike or we like it.

'*Appraisal' consists of an automatic association (conscious or unconscious) of an internal affective response/state either pleasant or unpleasant, either attractive or repulsive, etc., to the appraised stimulus or representation.*

It does not consist of a judgment of appropriateness or capability – possibly supported by additional justifications; on the contrary, it just consists of a subjective positive or negative experience/feeling associated with the stimulus or with the mental representation, usually previously conditioned to it in similar circumstances, and now retrieved.

This gives us a completely different 'philosophy' of valence and value: now the 'motto' is the other way around – in Spinoza's spirit – (see Figure 5.1)

'It is good/beautiful what we like/what is pleasant'

As a *cognitive evaluation of m is likely to give rise to some goal*: if the evaluator *X* believes something *m* to be a means for *X*'s goal *g*, *X* comes to have the goal 'of acquiring and using the means *m*' (we call *q* this instrumental goal); also *the emotional appraisal of m gives rise to a goal*: it activates a very general goal linked to the emotional reaction. This is the *cognitive* aspect of emotional appraisal. Positive appraisal activates an 'approach goal' ('to be close to *m*; to have *m*'), while negative appraisal activates a generic avoidance goal ('not to be close to *m*; to avoid *m* '). We consider *these sub-symbolic, implicit forms of 'evaluation' as evolutionary and developmental forerunners of cognitive evaluations*. Thus we believe the answer to the question 'do emotions imply evaluations?' depends on the level of analysis addressed.

In sum, in our view:

(1) *Appraisal is an associated, conditioned somatic response* that has a central component and involves pleasure/displeasure, attraction/repulsion. Here attraction/repulsion is not a

motor behavior but just the preliminary, central and preparatory part of a motor response. And pleasure/displeasure simply is the activation of neural centers.

(2) This associated response can be *merely central*, because the somatic-emotional component can also be reduced to its central trace (Damasio's somatic markers) and because emotions have a central response component which is fundamental. But of course this response can also be more complete involving peripheral overt motor or muscle responses or visceral emotional reactions.

(3) This associated response is *automatic, and frequently unconscious*.
- Appraisal is a way of *'feeling' something*, thanks to its somatic (although central) nature.
- Appraisal *gives 'valence'* to the stimulus because it makes it attractive or repulsive, good or bad, pleasant or disagreeable.
- Appraisal has *'intentionality'* i.e. the association/activation makes what we feel 'about' the stimulus, makes it nice or bad, fearful or attractive. It gives the stimulus the character that Wertheimer called 'physiognomic'. (How this happens, how the associated response is 'ascribed to', 'attributed to', and 'characterizes and colors' the stimulus; how it does not remain concurrent, but dissociated, is not so clear – at least to us – and probably just the effect of a neural mechanism).

(4) When it is a response just to the stimulus it is *very fast, primary*. It anticipates high level processing of the stimulus (like meaning retrieval) and even its recognition (it can be subliminal). In this sense the old Zajonc's slogan 'preferences need no inferences' proves to be right (although not exclusive: there are preferences which are based on reasoning and inferences; and also emotions based on this).

(5) There can be an analogous associative, conditioned, automatic *response to high level representations*: to beliefs, to hypothetical scenarios and decisions (Damasio, 1994), to mental images, to goals, etc.

We have to change our usual view of cognitive 'layers', where association and conditioning are only relative to stimuli and behaviors, not to cognitive explicit mental representations.

Any emotion as a response implies an appraisal in the above mentioned sense.

It implies the elicitation of a central affective response involving pleasure/displeasure, attraction/repulsion, and central somatic markers if not peripheral reactions and sensations. This is what gives emotions their 'felt' character. (While not all emotions presuppose or imply a cognitive evaluation of the circumstances).

5.5 Relationships Between Appraisal and Evaluation

Evaluation and affective appraisal have much in common: in particular, their function (Castelfranchi, 2009). Evaluations favor the acquisition of adequate means for one's goals, and the avoidance of useless or dangerous means, and precisely the same function can be attributed to emotions.

More than that: emotions – though they have traditionally been attributed the negative role of clouding and altering rational thought – seem to help at least some kind of reasoning. In fact, they provide 'non-conscious biases' that support processes of cognitive evaluation and

reasoning (Bechara, 1997), enabling one for instance to choose an advantageous alternative before being able to explicitly evaluate it as advantageous.[4]

However, all this should not prevent one from acknowledging the differences between emotional appraisal and cognitive evaluation, addressing the latter in their own right, and trying to establish their specific functions. For instance, in some context emotional appraisal by itself might prove insufficient for assuring adaptive responses, in that, the *more changeable and complex the world becomes* (because of the increasing number of goals and situations to deal with, and the complex relations among such goals and contexts), *the more one is in need of analytical and flexible judgments about objects and events*, rather than (or in addition to) more global and automatic reactions. In fact, evaluations allow one to make subtle distinctions between similar (but not identical) goals and means, and to find out the right means for some new goal, never pursued in the past.

Moreover, evaluations allow one to reason about means and goals, and to construct and transmit theories for explaining or predicting the outcome of behavior. Therefore, though emotional appraisal can be conceived of as an evolutionary forerunner to cognitive evaluation (as well as a valuable 'support' for it), being an evolutionary 'heir' does not imply maintaining the same nature as the forerunner; on the contrary, one might suppose that the same function has favored the development of different means, at different levels of complexity.

It is also important to consider that evaluation and appraisal about the same entity/event can *co-occur*, and give rise to *convergence* and enhancement of the valence, or to *conflicts*; in fact, either:

- the means that we are rationally considering for our ends are associated to previous or imagined positive experiences; or
- what I believe to be the right thing to do frightens me; what I believe to be wrong to do attracts me.[5]

Evaluation and appraisal can also derive one from the other.

It is possible to verbalize, to translate a merely affective reaction towards *m* into a declarative appreciation. This is for example what happens to the subjects in the experiment by Bargh and Chartrand, 1999. They do not realize that their evaluation is just a post-hoc rationalization of some arbitrary association (conditioning) they are not aware of.

[4] A number of studies conducted by Damasio and his collaborators (e.g. Bechara, Damasio, Damasio, and Anderson 1994) have pointed to the crucial role of emotion in cognitive evaluation and decision making. Their patients with lesions of the ventromedial prefrontal cortex show emotional blunting as well as difficulties in making decisions, especially in real-life contexts. If compared with normal subjects, they do not show stress reactions (as measured, for instance, by skin conductance response) when trying to make choices in uncertain and risky contexts (e.g. a gambling task). The interesting fact is that such emotional reactions, displayed by the normal subjects especially before making a wrong choice (i.e. a kind of choice previously associated with some punishment), help them to avoid it, and to opt for a less risky alternative. Such a choice is made before reasoning over the pertinent beliefs, including cognitive evaluations about the game, its options, and the possible strategies of decision making.

[5] On this, Damasio's model of the role of the somatic markers in decision-making looks rather simplistic: somatic markers do not 'prune' the tree of possible choices, but just add some weight or value to them; it is always possible that we decide to pursue a goal that actually disgust us, or that we do not pursue a goal that was very exciting and attracting.

It is also possible to go down the opposite path – from a cold evaluation to a hot appraisal; especially for personal, active, important goals, and in particular for felt kinds of goals like needs, desires, etc. (Castelfranchi, 1998).

This possible translation from one form to the other is very important, because it also helps to explain a very well known vicious and irrational circle of our emotional life (Elster, 1999). We mean the fact that we feel our emotional activation, what we feel towards *m*, as a possible evidence, *confirmation* of our beliefs that give rise to that emotion itself. So, for example, we start with a belief that *m* can be dangerous, we predict possible harm, on such a basis we feel some fear, and then this fear (as an appraisal of *m*) 'feeds back' on the beliefs and increases their certainty, i.e. confirms them; something like: 'Since I'm afraid, actually there is some danger here' (which is not such a rational evidence; it is a case of self-fulfilling prophecy and also the origin of 'motivated reasoning', (Kunda, 1990) (see Figure 5.1).

Applied to trust this means that there are two possible paths:

(i) On the one side, it is possible to formulate a judgment, and explicit evaluation ('John is honest; John is a serious guy; John is really expert in this') and then *feeling* a positive trust disposition towards John;
(ii) Feeling – for some unconscious and unclear analogy and evocation of previous affective experiences – a positive affective disposition of safety, reliability, other's benevolence, towards John, and on such a basis to formulate real beliefs and explicit evaluations on him.

In fact, why do we need to spend so much time on the theory of implicit, affective appraisal? Because there are forms of trust just based on this, not on explicit beliefs about *Y*. And also because trust usually has an affective component, is some sort of weak 'emotion', or at least a 'feeling', an affective disposition.

5.6 Trust as Feeling

Trust is also a 'feeling', something that the agent 'feels' towards another agent, something one 'inspires' to the others. It can be just *confidence* (similar to self-confidence) not a judgment. It can be not arguable and based on reasons or explicit experiences; it can be just 'dispositional' or just 'intuitive' and based on tacit knowledge and implicit learning.

At a primitive level (consider a baby) trust is something not express/ed/ible in words, not made of explicit beliefs about *Y*'s competence or reliability. It is a spontaneous, non reasonable or reasoned upon (non rational) reliance, and a feeling of confidence in a given environment or in a person.

What is this kind or this facet of trust?

Trust as a feeling is characterized by a sensation of 'letting oneself go', of relaxing, a sort of confident surrendering; there is an attenuation of the alert and defensive attitude (consider the trust/confidence of a baby towards her mother).

Affective components of trust result in a felt freedom from anxiety and worried; *X* feels safe or even protected; there is no suspicion or hostility towards *Y*, which is appraised/felt as benevolent and reliable ('S/he will take care of…'). Towards a benevolent *Y*, we are benevolent, good-willing; towards a good/skilled *Y*, we are not aroused, alerted, cautious,

worried (*X* could say: 'I am in your hands'; while *Y* would say 'Let yourself go, do not resist, *trust* me').

It is very important to note – coherently with our model of trust – that it is not true that in this kind of trust we (implicitly) evaluate/perceive only the good disposition, the *good will* of the other; we also appraise his *power* (ability, competence): for example, for a feeling of protection and safety, and for relying on the other as for being protected against bad guys with a sense of safety, also the perception of his physical strength and character is crucial.

Notice how these possible affective components of trust are coherent and compatible with our cognitive analysis of trust. However, they can also be independent of any judgment; they can be just the affective, dispositional consequence of an intuitive appraisal (cf. Section 5.2) and of learning. They can even be by default or just the result of lack of (bad) experiences, lack of negative evaluations (Chapter 4). In fact, a bad experience with *Y* would be a bad 'surprise' for *X*; something one was not suspecting at all. There are no really *subjectively* 'rational' justifications for that attitude, but it can be 'rational' in relation to a repeated experience or to evolutionary adaptive functions.

5.7 Trust Disposition as an Emotion and Trust Action as an Impulse

An emotion or feeling is 'caused' by, elicited by, it is spontaneously arousing on the basis of given beliefs (not only of 'stimuli', like for more simple and primitive emotions, like a very fast reaction of startle and fear due to a terrible noise, before any understanding of it). *The emotion is the 'response' of our mind-body to a given (mental) event (the internal configuration of our representation of the situation).* Those beliefs are not 'reasons' for the affective reaction, like they are 'reasons' for a believing or for a decision ('arguments').

This is our model of emotions in terms of 'cognitive anatomies', i.e. in terms of the specific beliefs and goals that are needed for eliciting and entertaining that emotion, which are necessary constituents, ingredients, but also bases for it. We would say that certain beliefs are the causes, the triggers of a feeling of shame, guilt, etc. and from an external point of view, they are also the 'reasons' for that emotion, but not from the internal, subjective point of view: emotions have no 'reasons' in the strict sense.

Feelings and emotions usually activate a goal (a behavioral tendency or a more abstract desire to be translated into actions). For example, fear activates the goal of escaping, of being safe; shame, the goal of disappearing; pity, the goal of being of help; guilt, the goals of repairing, atoning, and not doing the same again.

We can accept trust as feelings, for the same kind of analysis:

- a feeling, an affective response arousing from a given more or less explicit perception and appraisal of the world;
- an activated goal on the basis of this feeling and mental configuration.

We have also to remind our reader that in our model (as in several others), there is the possibility that the relation and path from assumptions to feeling can be reversed: instead of feeling fear because one sees or thinks that there is some danger, one can assume that there is some danger just because one is feeling fear; using the sensation as a sort of (non 'rational') 'evidence'.

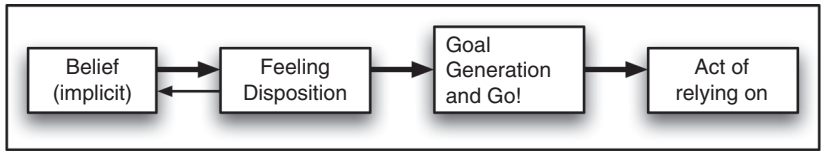

Figure 5.2 From Assumptions to Feeling

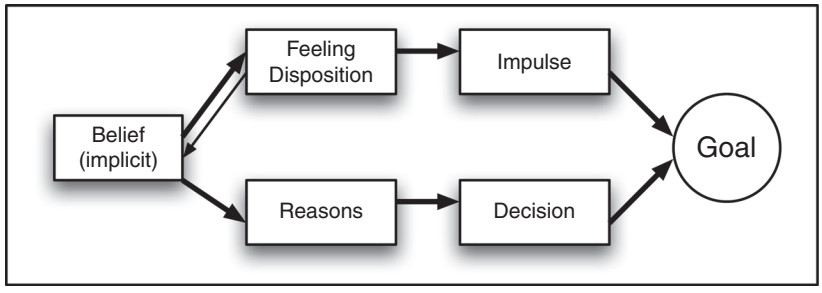

Figure 5.3 Two Kinds of Trust

So the model of the process would be something like that shown in Figure 5.2.

What matters here is the fact that a purposive behavior is generated not on the basis of a true 'decision' based on reasons, but is affectively activated, and has a strength not due to its supports and their credibility but just to the *intensity* of the affective arousal (Castelfranchi, 1998), and it not necessarily acted-out on the basis of a true 'deliberation' process.[6]

We might say that there are *two kinds of trust and two parallel paths of trust* (following in a sense a 'dual' model à la Kahneman) (see Figure 5.3).

In both cases, *trust* is not only the disposition towards the other (appraisal, evaluation, expectations), but is a winning goal of relying upon the other and the act of making oneself relied upon (many authors miss this point) and thus *making* oneself vulnerable and non-defended to the other.

One way or the other, complete trust arrives in both cases at a goal (the goal of counting on *Y* as for *g*) and an act (some sort of intention in action, even without any 'reason'), and then to a 'relation' which is not only evaluative, affective, or dispositional, but is actual, practical: of a concrete and exploited 'dependence' (relying/counting upon).

5.8 Basing Trust on the Emotions of the Other

The relationships between trust and affect/feelings are more complex. Not only can trust be just a feeling, an affective attitude and move towards *Y*; but affects from *Y*'s side are also very important.

[6] M. Tuomela's (Tuomela, 2003) analysis of trust is correct under this respect (as 'affective Trust not implying a true 'decision' and 'intention') but remains too restricted: only based on a perceived relation of mutual respect, of shared values, of recognition of my rights. This kind of trust can also be elicited just by love and protective attitude of the other towards me, or by pity, etc.

Y's perceived trustworthiness/reliability can just be based on the perception or ascription of affects and emotions in *Y*. We ascribe to the other not only goals and beliefs but also emotions and moods, and our positive (or negative) evaluation as for trusting *Y*, or our prediction can just be based on the perceived/ascribed emotion.

For example, I can rely on the fact that *Y* will carry out a given action (for example, helping a poor guy sleeping in the streets) because I see that he is really feeling pity or guilt, and he is pushed to care about that guy. Or I expect that *Y* will interpose between *Z* and me (*Z* is aggressing me) just because *Y* is very angry with *Z* and aggressively reacting to *Z*'s aggression. Or I expect help from *Y* because I realize that he has very nice sentiments and positive emotions towards me.

X's emotions and *Y*'s emotions can be strongly related with each other. On the one hand, *X*'s perception of *Y*'s emotional attitude can be based on *X*'s own emphatic emotions. On the other hand, *X*'s emotion of trust can be just the affective response and reciprocation of *Y*'s perceived positive emotions towards *X*. Moreover, *Y*'s perceived emotions can be elicited by *X*'s trustful and affective attitude (Chapter 8).

But again the affective-based and the reason-based attitude of trust can be (and usually are) mixed up with each other. Even an evaluation or a decision to trust *Y*, based on reasons and evidence, can elicit not only an affective disposition in *X* towards *Y*, but also an affective response in *Y*. Or vice versa, an affective disposition of *Y*, can be the base for a well-argued evaluation of *Y* as for a given task or relation (for example, as a babysitter).

5.9 The Possible Affective Base of 'Generalized Trust' and 'Trust Atmosphere'

We deny that the theory of generalized or diffuse trust (the so called 'social capital') is a separate and different model from the theory of interpersonal trust. In our view, the former must be built upon and derived from the latter (Chapter 6).

On the one hand, *trust capital is a macro, emerging phenomenon*; *but it must be understood also in terms of its micro foundations*.

On the other hand, conceptually speaking, the *notion of trust is just one and the same notion,*[7] at different levels of abstraction and generalization.

It is true that social trust relationships and networks (Chapter 9) are more than the simple summation of *X*'s trust towards *Y* (and *Z*, and *W*) and *Y*'s (etc.) trust towards *X* (etc.). Social trust builds richer 'relationships': *'I trust you; in so doing our relationship moves a little further forward.'* ((Holton, 1994) p. 68). But, why is it so?

Because *X*'s trust is also a *signal*, it exhibits an internal evaluation, etc.; because it gives or shows to the other the opportunity to harm *X* and thus *the opportunity to choose not to harm X*, of favoring *X*; because it elicits gratitude, relax, trust improvement; because it creates reciprocity of both attitude and behavior.

5.10 Layers and Paths

In conclusion, *in our model there are three layers or stages and two paths or families of trust* (see Figure 5.4).

[7] Both, as emotional disposition, feeling, and action tendency; and as an explicit evaluation, belief, and grounded expectation and decision.

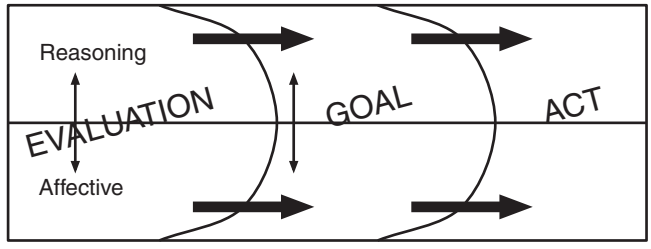

Figure 5.4 Layers and Paths of Trust

When discussing the layers we have:

1) Appraisal and disposition towards Y;
2) Goal generation and selection;
3) Action and the consequent practical relation (*counting on*).

About the paths/families we have:

1) Reason-based trust; and
2) Affective or feeling-based trust.

In Figure 5.4, the phase of evaluation corresponds with the double potential question: What I think/feel about Y (as for my depending for τ).

The two roots are not completely alternative and independent: they can coexist and influence each other. This is why we put some crossing arrow between them; in order to signal possible influences of the feeling on the goal, although this is part of a real deliberation, and is based on explicit reasons and evaluations.

As we have already said, in this book we are mainly focused on reasons and decision-based trust, not on the merely affective and 'impulsive' one; but they are 'equifinal' (respond to similar 'functions') and the former is an evolved form of the latter.

5.11 Conclusions About Trust and Emotions

Here we are just analyzing an affective-based form of trust, not the relationships between trust and emotions in general. However, notice that our socio-cognitive analysis of trust as psychological state and behavioral attitude in terms of cognitive and motivational components, provides a systematic and grounded basis for predicting and explaining the relationships between trust and emotions.

On the one hand, some trust constituents are also constituents of emotional reaction, which – thus – can co-occur with trust. This is the relationship between trust and hope, or between trust and worries/fear; or trust in Y and sympathy/benevolence towards Y; or trust and relaxation, not to feel exposed or in danger with Y.

Moreover, the trust attitude and decision exposes to new affective states and reactions, like 'disappointment' (since trust implies positive expectations), or to 'feel betrayed' (when trust relies on *Y*'s adoption and commitment).

A good componential analysis (cognitive-motivational anatomy) of a mental state should be able to account for the relationship between different kinds and levels of that phenomenon, of the relationships between it and close or possibly co-occurring mental states, and conceptual families.

We do not propose to provide a systematic treatment of that here, but in different parts of this book we stress the justified relations between trust and other affective or cognitive states: expectations, evaluations, uncertainty, anxiety, fear, surprise, hope, faith, optimism, sympathy, and so on.

References

Arnold M.B. (1969) *The Nature of Emotion.* Penguin Books.

Bargh, J. A., and Chartrand, T. L. The unbearable automaticity of being. *American Psychologist*, 54: 462–479, 1999.

Bechara, A., Damasio, A.R., Damasio, H. and Anderson, S.W. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50: 7–15, 1994.

Bechara, A., Damasio, H., Tranel, D. and Damasio, A.R. Deciding advantageously before knowing the advantageous strategy. *Science*, 275: 1293–1295, 1997.

Castelfranchi, C. (1998) To believe and to feel: The case of 'needs'. In D. Canamero (ed.) *Proceedings of AAAI Fall Symposium 'Emotional and Intelligent: The Tangled Knot of Cognition'*, AAAI Press, 55–60.

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*. Heidelberg, Springer, LNAI 1814, 76–106.

Castelfranchi, C., Miceli, M. (2009) The cognitive-motivational compound of emotional experience. *Emotion Review*, 1 (3): 221–228.

Damasio, A. R. (1994) *Descartes' Error*. New York, Putnam's Sons.

Elster, J. (1999) *Alchemies of the Mind: Rationality and the Emotions*, Cambridge University, Press Cambridge UK.

Frijda, N. H. (1986) *The Emotions*, Cambridge University Press.

Frijda, N. H. (1988) *Cognitive Perspectives on Emotion and Motivation*, Springer.

Holton, R. Deciding to trust, coming to believe. *Australian Journal of Philosophy* 72: 63–76, 1994.

Kunda, Z. (1990) *The Case for Motivated Reasoning*. Princeton University.

Lazarus, R. S., Averill, J. R., and Opton, E. M. (1970) Towards a cognitive theory of emotion. In M. B. Arnold (ed.), *Feelings and emotions, The Loyola Symposium*. New York: Academic Press. (pp. 207–232).

Mandler, G. (1975) *Mind and Emotions*, New York: Wiley.

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.

Ortony, A., Clore, G. L., Collins, A. (1990) *The Cognitive Structure of Emotions*.

Scherer, K. R. (1986) *Experiencing Emotion*. Cambridge University Press.

Scherer, K. R. (1999) Appraisal theories. In T. Dalgleish, and M. Power (eds.). *Handbook of Cognition and Emotion* (pp. 637–663). Chichester: Wiley.

Tuomela, M. (2003) A Collective's Rational Trust in a Collective's Action. In Understanding the Social II: Philosophy of Sociality, *Protosociology*. 18–19: 87–126.