Chapter 6
The irritation of doubt: When is it OK to scratch your beliefs?

Fabio Paglieri

Abstract Building upon Peirce's notion of "the irritation of doubt", I outline the role doubts play in belief formation, drawing a distinction between two different phenomena: Peircean doubts and scary doubts. I argue that the former are an essential and useful ingredient in our belief dynamics, whereas the latter are potentially dangerous and yet unavoidable biases, even though they might occasionally be redeemed by their practical merits. This analysis is intertwined with Castelfranchi's views on belief formation and goal processing, with the aim of highlighting a common root between pragmatism and goal theory, and to provide a unifying picture of the mind as a goal-directed, coherence-seeking control system.

1 Introduction

In one of his many illuminating passages, Peirce observed that «the action of thought is excited by the irritation of doubt, and ceases when belief is attained; so that the production of belief is the sole function of thought» (1958, p. 118).[1] Nowadays "thought" is called cognition and is no longer conceived as pertaining solely or even mainly to the production of belief. Nevertheless, there is still much to be learned from Peirce's remark, as well as good chances of misunderstanding its meaning. In what follows I try to elucidate both Peirce's insight and some misconception that might plague it, and to stress its connections with Cristiano Castelfranchi's own views on belief dynamics and, to a minor extent, goal dynamics. This will serve not only to build a bridge between

Fabio Paglieri
Goal-Oriented Agents Lab (GOAL), Institute of Cognitive Sciences and Technologies, CNR, Italy
e-mail: fabio.paglieri@istc.cnr.it

[1] Originally published as "How to make our ideas clear" in Popular Science Monthly 12 (January 1878), pp. 286-302.

the father of semiotics and the father of goal theory, but also to formulate a general hypothesis on the main function of the mind.

So, what kind of doubt works as an irritant for cognition, according to Peirce? As it has been noted before (Thagard 2004; Magnani 2009; Magnani and Bertolotti 2011), and as it is clear from the rest of Peirce's essay, he is not referring to some kind of Hamletic doubt on ontological or moral dilemmas ("To be or not to be?"), nor to the all-encompassing Skeptical doubt, but rather to mundane doubts concerning practical matters: it is the kind of doubt we encounter when we are presented with any ambiguous or unclear situation that requires us to form an opinion[2] on the facts in order to act accordingly and, if we are right, adequately. It is the doubt we face when we wonder about the nature of tonight's party to decide what to wear, or about the likelihood of rain in the afternoon to choose whether to take the umbrella in the morning, or about the meaning of our boss' remark to find a suitable response, or about the relative merits and shortcomings of various candidates in an election to establish how to vote. These are the kind of doubts that, according to Peirce, prompt our cognition to operate and move towards a certain state of belief. Let us call them *Peircean doubts*, for the sake of simplicity.

Now, it should be clear that Peircean doubts are an irritant insofar as they are impractical: the subject needs to know what to do, and for that s/he must move beyond the current state of doubt and reach some belief on the issue at hand, no matter how temporary. It should also be clear, though, that there is another type of doubts, which are not those contemplated by Peirce here, and that these doubts can also work as an irritant, but in a very different (and stronger) sense. I am referring to those doubts that threaten one's self image and/or social image, such as doubting one's professional or relational adequacy, moral fiber, religious faith, ethical values, personal engagements with family and friends, and the like. Let us call them *scary doubts*, because that is what they do: they scare the person into believing something, often (not always) with little regard for evidence, because the persistence of doubt would be painful or otherwise distasteful for the subject.

This distinction is relevant because Peircean doubts are not scary doubts, and their respective effects on cognition should not be confused. Yet it is easy to confuse them, since both types of doubt pressure the mind towards belief: but they do so in different ways, for different reasons, and serving different functions. While the exact differences between Peircean doubts and scary doubts will be the topic of the next sections, I want to anticipate their crucial difference. The influence of Peircean doubts is a fact of life, and Peirce was absolutely right to remind us that our strife for belief is essentially a response to the (practical) irritation they cause to our cognitive system: we need to reach a somewhat stable assessment of reality, even if precarious and fallible, in order to operate effectively, so it is unsurprising and certainly adaptive that our cognition evolved to answer that need. But the influence of scary doubts on belief formation and change (or lack thereof) is something that needs to be proven, and indeed I will review some evidence about it. Moreover,

[2] To Peirce, famously, this process was mainly one of abduction: for present purposes, however, it is unimportant what kind of inferential mechanism is involved in belief formation.

even after establishing that some of our beliefs are born out of fear rather than reason, this does not imply at all that fear-induced belief formation is a "good thing", either morally, rationally, or adaptively. Arguments are needed for each of these claims, and in what follows I will try to assess their merits and limits. For the time being, let us just emphasize that Peirce's thesis on the role of doubt in prompting cognition towards belief was intended for (and makes sense of) practical doubts on mundane matters, not scary doubts on issues of personal relevance.

2 Peircean doubts, epistemic bubbles and degrees of belief

Let us define a Peircean doubt as follows:

PROPOSITION 1 (Peircean doubt) *An agent X has a Peircean doubt on the issue P when (i) X is undecided regarding whether P or not-P is the case, and (ii) X needs to reach a temporary conclusion on P in order to act in the pursue of X's goals.*

Most emphatically,

COROLLARY 1 *It is not required that X has anything of importance at stake in resolving the current state of doubt one way or another, aside from being able to act.*

Peircean doubts are minor inconveniences that we encounter in the ongoing transition from ideation to realization of our daily deeds: for a variety of reasons (ignorance, uncertainty, forgetfulness, perceptual limits, etc.), we find ourselves with some gaps in the list of things we need to command in order to plan a certain action, and these gaps have to be fixed immediately and without much conscious effort or reflection. Repairing Peircean doubts typically does not involve any sophisticated reasoning task: agents take them in stride and simply assign a value to the previously undecided item, based on other relevant knowledge and (if pertinent) past experience. Interestingly, the agent treats the solution to a Peircean doubt as something well established (well enough to warrant acting upon it), and yet has no reservation in reconsidering it – that is, the kind of belief prompted by the irritation of Peircean doubts is not particularly resistant to update and change.

As a case in point, imagine that Luigi goes to the train station in Rome with the aim of taking the 11:00 am express train to Naples: upon entering the station at 10:45 am, Luigi looks up at the electronic board listing all departing trains and their platforms. To his dismay, the 11:00 am is listed as on time, but no platform is yet indicated, leaving Luigi in the dark on where to go (a typical instance of Peircean doubt). Since Luigi frequently takes this train from this station, he happens to know that (i) the train typically leaves from platform 20B, and (ii) the electronic board is often defective and does not show the platform until the very last moment. Without hesitation, Luigi solves his Peircean doubt by assuming that today the train will leave from platform 20B, and thus proceeds in that direction. At this point, Luigi is quite firm in his conviction that he is moving in the right direction, and he is certainly willing to act upon this notion. But clearly this would not preclude him from changing his mind in the blink of an eye, should further considerations come into play – e.g., an audio announcement informing passengers that the 11:00 am train to Naples is leaving from platform 7C. Even more importantly, we find nothing to object in Luigi's beliefs and actions, and in fact there would be nothing to fault in his reasoning even if the right platform number turned out not to be 20B.

As the example shows, there are several typical features of Peircean doubts and of the beliefs originated by them:

1. Peircean doubts need to be solved as a matter of practical necessity;
2. solving them does not serve to alleviate any special concern of the agent, but merely to enable some practical course of action;
3. the resulting belief is sufficient to warrant acting upon it – in fact, warranting action is the reason why the irritation of Peircean doubt has to be satisfied;
4. the resulting belief is perfectly defeasible, and the agent has no special resistance to change it, if given reasons to do so.

Considering the strong emphasis placed on practical reasoning and action execution in this analysis of Peircean doubts, one might be tempted to suggest that perhaps belief is not the right mental attitude to invoke as their solution – contra Peirce's original suggestion. If all that really matters in these cases is to take as granted something (temporarily) in order to act, then it would seem that these are instances of *acceptance* rather than belief, in the sense proposed, among others, by Cohen (1989) and Bratman (1992). Elsewhere (Paglieri 2009) I have argued that this pragmatic notion of acceptance is not especially helpful, since it is redundant with respect to explanations based on beliefs and goals: whenever it is tempting to say that the agent is in a state of acceptance of *p* (in our example, accepting that the train will leave on platform 20B), it is also the case that the agent's actions can be as easily (and more parsimoniously) explained in terms of doxastic and motivational attitudes – in Luigi's case, it would be natural to say that he wants to catch the train and, based on his past knowledge of Rome train station, he believes running to platform 20B is the most sensible course of action. So making too much of pragmatic acceptance does not grant us any special insight on the agent's cognitive practices – or so I claim (for further arguments and details, see Paglieri 2009).

Another tempting equivocation of Peircean doubts is the idea that our tendency to quench their irritation creates a special *vulnerability* to our doxastic practices, making us prone to believe much more than we should, given extant evidence. This interpretation would become especially pernicious if paired with John Woods' notion of *epistemic bubbles* (2005), that is, the observation that knowledge and belief are, in many cases, indistinguishable from a first person perspective. If Peircean doubts made us prone to generate hasty beliefs, and if epistemic bubbles made us mistake such beliefs for knowledge, then our predicament would seem dire indeed. Fortunately, neither Peircean doubts nor epistemic bubbles have anything to do with any special vulnerability of our cognitive practices. Let us see why.

Woods' reflections on epistemic bubbles stems from an observation on the phenomenology of knowledge: whenever we think we know that *p*, this state is indistinguishable for us from the state of actually knowing that *p*. There is nothing to tell apart the experience of knowing that *p* from the experience of thinking to know that *p*: in fact, these two experiences are identical, from a first-person perspective. Of course, from a third-person perspective it is perfectly clear (with variations, depending on your favored epistemology) what is that distinguishes thinking to know that *p* from actually knowing that *p*: but a third-person perspective is precisely what we cannot have concerning our own knowledge states, and the point Woods is making is that, no matter your preferred epistemology, you will remain blind on whether you actually know that *p* or you just think you know it. In his apt metaphor, we are all caught in an epistemic bubble, defined as follows:

> A cognitive agent *X* occupies an epistemic bubble precisely when he is unable to command the distinction between his thinking that he knows *P* and his knowing *P*. Corollary: When in an epistemic bubble, cognitive agents always resolve the tension between their thinking that they know *P* and their knowing *P* in favour of knowing that *P* (2005, p. 740).

A careless reader might interpret the epistemic bubble thesis as simply saying that people often mistake beliefs for knowledge, thus overestimating the quality of their doxastic states. But nothing of the sort is implied here. First of all, the thesis is restricted to those beliefs that are experienced as knowledge by the agent (as in "I know the bus will arrive", "I know what is best for my children", "I know you cannot mean harm to me", and the like), and it is not meant to be true for all senses of the word "belief" – most notably, it is not meant to be true for (weak) gradual beliefs, suspicions, hunches, and the like (on this point, see Woods 2005, p. 744).[3] Secondly, experiencing something as being known does not necessarily imply being entrenched in this conviction, according to Woods (and I agree): it is a common experience to take something as being known one moment, and then abandon it without regret as soon as its inadequacy becomes apparent. That is, finding out that something we thought to know was actually mistaken (hence not known at all) is no big surprise for any minimally self-reflective agent, and does not typically cause any major discomfort. In fact, we are so jaded with respect to the volatility of our claims to knowledge precisely because we are accustomed to living in an epistemic bubble, and we have learned to tolerate well our incapacity to tell apart true knowledge from the mere appearance of it.

[3] Actually, Woods would probably go further and claim that many things often labeled as beliefs in the literature do not deserve that label, and would instead be better classified as other kinds of mental attitudes, e.g. suspecting that p, being willing to bet on p, finding plausible that p, and so on (2005, p. 744-745). This view is consistent with Bas van Fraassen's invitation to endorse a more liberal descriptive epistemology, one that admits of greater variety than just belief, disbelief or neutrality (2001, p. 165). While I agree that we should remain open-minded on what mental states to include in our epistemology, I also think that Ockham's razor should shave our liberal inclinations here, and that mental attitudes should not be multiplied without cause. Since using "belief" in its more general sense has no significant repercussion for current purposes, here I will stick to this traditional usage.

The crucial point is that the subjective strength of our convictions does not depend on them being experienced as knowledge rather than belief – contra the epistemologist's prejudice according to which knowledge is something more/better than belief (e.g., true justified belief, in the most widespread recipe). Woods' notion of epistemic bubbles is not committed to such prejudice: it is a thesis on the fact that the distinction between belief and knowledge, perfectly intelligible in a third-person perspective, is immaterial from a first-person perspective, but it does not imply that the agent will become fixated on whatever beliefs s/he misattributes as knowledge.

So, what is the right way of connecting Peircean doubts with epistemic bubbles, if any? There is a thin but clear link: it could very well be the case that our tendency to take belief as knowledge is inspired by the practical need for a partially stable (and quick) assessment of reality, as emphasized by Peircean doubts. Second-guessing what we think to know is, in general, the wrong way of reacting to a Peircean doubt: whether that knowledge will turn out to be correct or not is relevant only to the extent that the resulting action will end up being successful – and if it does, that is all that matters to the agent. This is why issues of justification (in the sense of having a good case for *p*, not just a cause for believing *p*; see Woods 2005, pp. 759-761) becomes typically irrelevant to consider, or at least secondary.

But it is certainly *not* the case that Peircean doubts plus epistemic bubbles imply belief rigidity. This is not the case because knowledge does not differ from belief in terms of doubt-resolution: believing that *p* solves Peircean doubts about *p* as effectively as knowing that *p*, from a first-person perspective (precisely due to our epistemic embubblement). The difference between belief and knowledge is in terms of objective truth and valid justification, and it is a difference that, according to Woods, we have trouble grasping in a first-person perspective. So Peircean doubts and epistemic bubbles can happily co-exist, without engendering any special risk of tunnel vision or cognitive fixation. For the same reasons, Peircean doubts and epistemic bubbles do not (and are not meant to) offer an explanation to extant cases of unreasonable resistance to changes in one's beliefs. For a diagnosis of such cases, one has to look at scary doubts and their impact on belief formation: this will be the business of the next section.

Before moving on to that, however, it is interesting to put Peircean doubts in contact with the idea that, in believing that *p*, extant evidence on *p* produces a lower and an upper boundary for our *degree* of belief in *p*, and the interval between such boundaries represents the agent's ignorance on *p* – in the vein of Dempster (1967) and Shafer (1976). The interesting psychological twist in looking at beliefs that way is that whatever assignment of confidence *x* to *p* within *p*'s ignorance interval is compatible with the existing evidence, hence legitimate. In other words, it is not a matter of evidence whether the agent believes more or less strongly that *p*, within the current ignorance interval for *p*. What is the import of that for Peircean doubts? Simply put, it allows us to speculate the following:

PROPOSITION 2 (Peircean confidence) *When faced by a Peircean doubt on p and with an ignorance interval for p equal to [A, Z], solve the doubt by setting the degree of belief in p to the upper boundary Z.*

Let us see how Peircean confidence might work in our train example. We will stipulate that Luigi's past experience in taking the 11:00 am train to Naples from Rome provides a lower boundary for the belief "The train will leave from platform 20B" equal to 0.5: that is, current evidence does not justify being more skeptical than that on platform 20B being the right one. On the other hand, Luigi did experience unexpected last-minute changes of platform in the past, so let us say that his upper boundary for that belief is equal to 0.9: that is, current evidence does not justify being more certain than that about the rightness of that belief. Peircean confidence predicts that Luigi will automatically set his belief to the highest justified degree – in this case, 0.9. Interestingly, this view explains both the fact that the agent is ready to act on beliefs prompted by Peircean doubts (unless there is strong evidence against them, i.e. the upper boundary is too low to warrant action), the fact that doing so is perceived as fully rational (as it is, in the sense of being within the boundaries of what is permissible, given current evidence), and the fact that even discovering such beliefs to be false is not so shocking, since they were conceived from the start as an informed guess under conditions of partial ignorance. Indeed, a subsequent change of belief would be surprising for the agent only if it revealed the original boundaries to be mistaken: for instance, if Luigi discovered from a train assistant that the 11:00 am train to Naples in fact *never* or *rarely* leaves from platform 20B, this would give him pause, and rightly so. But the conflict here would not be with the belief that today the train is leaving from platform 20B, but rather with the past evidence that Luigi used to reach that belief. And Luigi would have solid reasons to be puzzled, because in this case either the new information is mistaken, or what he thought was good evidence in fact wasn't.

Even if Peircean confidence works well in this example, it might be defective as a generalization of how we react to Peircean doubts. For instance, one might object that Peircean confidence, once applied to issues on which there is no evidence for or against, delivers the counterintuitive verdict that we should have full belief in such matters, since their ignorance interval is [0, 1]. To this, there are two answers: first, complete ignorance on an issue relevant for Peircean doubt occurs more rarely than one might expect; secondly, it is reasonable to assume that, when faced with contradictory beliefs, one will endorse only those with the higher degree, and none at all if they are all evenly matched. Under this assumption, full credence for states we are ignorant about becomes unproblematic, as we shall see.

Let us start from the first point, using again Luigi's predicament as a case in point. A critic of Peircean confidence may be tempted to reason as follows: Luigi is completely ignorant on whether the 11:00 am train to Naples may leave from any platform *different* from 20B, hence he should believe (for Peircean confidence) that the train will leave from any of those platforms with degree 1, which is greater than 0.9 – hence Luigi should prefer to run for any platform, except the one he knows it is likely the train will leave from. This is paradoxical, thus Peircean confidence is mistaken as a general criterion. However, this argument is clearly flawed, because Luigi is not at all ignorant on whether the train will leave from any of the other platforms: the very fact that he has evidence of platform 20B being the right one is evidence *against* the train leaving from another platform; conversely, whatever evidence against the train leaving from platform 20B is evidence *in favor of* the train leaving from somewhere else. In short, the belief "The train will leave from platform X", with X being any platform other than 20B, in this example has the ignorance interval [0.1, 0.5], hence it is to be believed, by Peircean confidence, with degree 0.5, which is lower than 0.9 and explains why Luigi runs towards platform 20B and nowhere else.

The second point is more delicate. Imagine an issue *p* for which there is indeed full ignorance, and yet such as it raises a Peircean doubt – for instance, you are considering whether to buy at a certain price a toy for your niece, and you do not have any clue on whether that price is appropriate or not.[4] Now, in this case your ignorance interval on "The price is right" is [0, 1], so Peircean confidence would have you giving full credence to it, hence buying the toy without a moment of hesitation – which does not seem the right option, given your utter ignorance on the matter (after all, the toy may be outrageously over-priced, for all you know). But on further reflection, it is obvious that also the ignorance interval for the contrary belief "The price is *not* right" is [0, 1], hence Peircean confidence would command full credence for this belief too. But you cannot have it both ways, hence Peircean confidence leaves you with no clue on what to do – which is precisely what should happen, assuming you are indeed totally ignorant on the matter. That is, Peircean doubts under conditions of complete ignorance are *not* meant to be solved.

To see why, consider the following variation of Luigi's predicament: imagine now that Luigi has never taken the 11:00 am train to Naples from Rome and has no prior information on what platform the train is supposed to leave from – he is in a state of full ignorance on this issue. Upon looking at the electronic board and realizing that no platform is indicated there, what should Luigi do, to behave rationally? Even more crucially, what would *you* do, if you were in that situation with a pressing need to catch that train? You would certainly not start running towards a random platform, nor should Luigi. What you would do is to accept that your Peircean doubt on the train's platform currently has no satisfactory solution, and thus proceed to look for more information – either waiting in front of the electronic board, hoping the platform number will soon appear, or going in search of someone who might point you in the right direction. This is what any reasonable person would (and should) do in a state of total ignorance, and this is why it is OK for Peircean confidence to leave you in the dark on what to believe in such cases. Otherwise, Peircean confidence

[4] Once again, it is worth emphasizing how unnatural the assumption of total ignorance is, in matters of Peircean doubts: even if you did not enter a toy shop in your whole life and were never exposed to the prices of toys (which is unlikely), you would still have a lot of background experience on what is the value of commodities in general. You would know, for instance, that a price of 1 billion dollars is excessive for a doll, while a price of 1 cent is unrealistically cheap – comparisons with the price of other goods would be enough to give you evidence for such rough estimates.

would in fact become *overconfidence*, which is not a reasonable reaction to Peircean doubts.

In light of these considerations, it is possible to propose a revised version of Peircean confidence, to better accommodate cases of complete ignorance:

PROPOSITION 3 (Peircean confidence revised) *When faced by a Peircean doubt on p, consider all the candidate beliefs* p^i *that would solve that doubt, each of them with a respective ignorance interval* $[A^i, Z^i]^i$. *Then take only the candidate (or the internally consistent subset of candidates) that has the highest upper boundary, and believe it with a degree equal to such upper boundary. If there are two or more conflicting candidates that share the same upper boundary, do not believe anything at all on the issue at hand and look for more information.*

The morale of these reflections is that Peircean doubts are not solved by any sleight of hands, such that states of complete ignorance can be magically turned into full-blown beliefs, while retaining a valid claim to rationality. What happens is both more mundane and more healthy: Peircean doubts are solved, whenever possible (to wit, in the absence of total ignorance), by taking a carefully optimistic outlook on the evidence we already possess and acting on it.[5] This is both *optimistic*, insofar as we pick the upper boundary of the current ignorance interval, and *careful*, because that boundary is still within what is rational to surmise in light of present evidence. Importantly, this way of forming defeasible beliefs as a reaction to the irritation of Peircean doubts is both automatic and effortless – hence the simplicity of the principle of Peircean confidence.

3 Scary doubts, defensive beliefs and biases

In order to emphasize their difference with Peircean doubts, let us define scary doubts as follows:

[5] It may seem natural to think that this carefully optimistic stance is modulated by personal inclinations and contextual factors. After all, a very prudent person would be naturally more reserved in giving credence to things that are uncertain, while special circumstances (e.g. urgent matters where one lacks better alternatives) may justify acting on wild guesses, even beyond the upper boundary of one's ignorance interval. However, I think it is more natural and more parsimonious to think that these effects concern what we do with our beliefs, rather than how we form them. Imagine the train scenario happens to Luisa instead of Luigi, and let us say that Luisa is much more prudent than Luigi: as a result, she does not rush to platform 20B, but rather waits until the latest possible moment to see what number will appear on the electronic board. I see no reason to claim that Luisa is any less confident than Luigi on whether the train will in fact leave from platform 20B, assuming they have the same evidence: the point is rather that she does not want to take the small risk of running to the wrong platform, while Luigi accepts it. Even more clearly, the fact that on occasion acting on a wild guess may be the rational (or even only) option has nothing to do with the doxastic status of that guess: that is, we do not become any more convinced of a wild guess just because we decide or are forced to act upon it.

PROPOSITION 4 (Scary doubt) *An agent X has a scary doubt on the issue P when (i) X is undecided regarding whether P or not-P is the case, and (ii) at least one of these options would, if believed, undermine some important goals of X.*

A scary doubt has no neutral solution for the subject: when I wonder whether or not my wife loves me, how I answer this doubt deeply affects my well-being, besides and beyond how it affects my conduct. More generally, with scary doubts at least one[6] of the candidates for belief has a negative impact on the agent's goals – more precisely, on some *important* goal of the agent. This is very different from what happens with Peircean doubts, where solving the doubt is a matter of practical necessity, but nothing relevant for the agent is at stake in solving it one way or another (provided such solution turns out to be pragmatically right).

Scary doubts entertain important relations with a special class of beliefs, defined as follows:

PROPOSITION 5 (Defensive belief) *An agent X has a defensive belief p if believing p avoids undermining an important goal of X and prevents X from being in a state of scary doubt on p.*[7]

In order to avert any equivocation, let us first clarify that a belief can be defensive even if it is held on the ground of legitimate evidence, and not because of whatever scary doubt it helps keeping at bay. To get back to the previous example, the belief that my wife loves me is a defensive belief, with respect to my goal of being loved by her, even if I have perfectly sound reasons to endorse it: her daily manifestations of devotion, our mutual happiness, the harmony of our family, and so on. Dropping that belief would hurt me, while maintaining it prevents that from happening: this fact is independent from whatever other reason I might have to hold that belief.

In other words, defensive beliefs are a broader category than *fear-induced beliefs*, defined as follows:

PROPOSITION 6 (Fear-induced belief) *An agent X has a fear-induced belief p if p is a defensive belief which was formed and/or is maintained, partially or solely, to avoid undermining an important goal of X and to prevent X from being in a state of scary doubt on p.*

The distinction between defensive beliefs and their sub-class of fear-induced beliefs captures the fact that not all defensive beliefs are irrational, in the sense of lacking proper reasons for being held. Conversely, having proper epistemic backing does not make a defensive belief any less defensive, since what makes

[6] The "at least" specification is needed to cover also cases where many or even all answers to the doubt are damaging for the agent: e.g., imagine a 50-years-old unemployed man who never had a sentimental partner in his life, seriously wondering whether his misfortunes indicate lack of character or some congenital cognitive deficit. We might label these extreme instances as "scary dilemmas", and treat them as severe instances of scary doubts.

[7] Defensive beliefs are only a type of self-serving beliefs: another obvious candidate are fulfilling beliefs, that is, beliefs that actually satisfy some important goal of the agent. Since here I am interested in scary doubts, I will confine my analysis to defensive beliefs, even though many considerations relevant for them apply also to fulfilling beliefs.

it so is a relationship with the agent's goals, not with evidence. Regardless, the presence of scary doubts create an influence on belief formation that is very different from the pressure towards belief countenanced by Peircean doubts.

Before discussing the specifics of such influence, there is an interesting question worth considering, albeit only cursorily: Is it better to live in a state of scary doubt or to face the worst possible outcome? More precisely: Is a scary doubt more or less scarier than the conclusion we fear to draw from it? The question is not idle, as the wife example immediately reveals: it is not at all clear that a man constantly wondering about his wife's affection is better off than another man who firmly believes his wife to be no longer in love with him. After all, we all have experiences (both first-hand and reported) on how sometimes fearing something is much worse than facing it – or, as the saying goes, "the devil is not so black as he is painted". The same could well apply to scary doubts: it is perfectly possible that persisting in a state of scary doubt is no better, subjectively speaking, than reaching a negative conclusion on the matter in doubt. However, this does not change the basic relation between scary doubts and defensive beliefs, since the latter solve both the uncertainty *and* the fear of the worst. My belief in the love of my wife prevents me both from worrying about it and from believing that she doesn't.

But is there any evidence that scary doubts, so defined, have an influence on our beliefs? Indeed, social psychology abounds of such evidence, and entire theories have been developed to study and articulate similar effects: in different guises and with different emphasis, cognitive dissonance (Festinger 1957; Aronson 1969), confirmation bias (Wason 1960), motivated reasoning (Kunda 1990), self-verification (Swann and Read 1981; Swann 1999), self-enhancement (Alicke and Sedikides 2011), self-perception (Bem 1967), adaptive preference formation and "sour grapes" effects (Elster 1983), and good old fashioned wishful thinking (Greenwald 1980; Taylor and Brown 1988), they all deal with various aspects of the influence of scary doubts on belief formation. The strength and the scariness of such doubts may vary, from the relatively severe (e.g. in self-verification) to the comparatively mild (e.g. in cognitive dissonance), yet all these approaches have in common the basic view that beliefs can be shaped by a pressure to avoid undesirable consequences – that is, a pressure to avoid jeopardizing one's goals by coming to believe something we would not like. Moreover, in all these contexts, albeit with variations, it is assumed that this influence is partially independent from, and potentially in contrast with, evidence-based reasons to hold or not a given belief.

My point here is not to review this large literature, but rather to mention it as proof of the extent to which scary doubts influence belief formation, and also of the difference between such influence and the impact of Peircean doubts on our beliefs. In a nutshell, the key distinction is in that scary doubts pressure us to reach a *specific* conclusion on the matter under consideration, whereas Peircean doubts only demand that we reach *some* conclusion, with no constraint on which one it should be. This difference is capital, since it defines scary doubts as *biases*, whereas Peircean doubts are revealed as mere expressions of a *practical need* for information to act upon. Granted, scary doubts may induce, on occasion or even typically, very beneficial biases, as I shall discuss in a moment. But this does not change the fact that their effects are biases, insofar as they systematically pressure belief formation in a direction unrelated with considerations of accuracy or evidence.[8]

Also the converse is true, of course: not all biases are necessarily liabilities for our cognitive processes – indeed, some biases can be highly advantageous. This point is often missed in cognitive psychology, where the terms "bias" carries an indisputable negative connotation. For instance, much of the current tension between the heuristics and biases approach of Kahneman and Tversky (Kahneman et al. 1982) and the adaptive toolbox idea championed by Gigerenzer (Gigerenzer et al. 1999) hinges on considering biases either a negative or a positive asset for cognition. However, it is possible (and I believe desirable) to use a neutral notion of bias, as a systematic influence over a certain disposition which is independent from the typical and/or normatively correct reasons to hold such disposition – in the case of belief, these typical/correct reasons would roughly amount to evidence-sensitivity and internal consistency. All influences that are independent from such considerations and yet typically affect belief formation can legitimately be defined as biases, with no evaluative judgment attached to this label. In fact, such judgment can only be applied meaningfully to specific biases or, more precisely, to applications of certain biases in specific contexts.

This is true also for the kind of defensive biases discussed here, and in particular for fear-induced beliefs. The general point is rather obvious: defending something can be good or bad depending on *what* you defend, and *why*. For instance, tampering with one's own beliefs in order to avoid impairing the capacity to function effectively (e.g. due to a depressive crisis induced by low self-esteem) may be a very wise thing to do. More generally and less dramatically, the fact that our goals have a biasing influence on our beliefs can have beneficial effects on our behavior even when no psychopathology is looming over us, e.g. in terms of coherence of conduct. As argued in details by Miceli and Castelfranchi (2012), a more or less acute obsession with preserving a coherent self-image is a powerful motive to avoid excessive deviations from one's typical conduct, thus ensuring greater behavioral stability: whether or not this turns out to be a good thing depends on whether a given individual is set upon a virtuous or vicious path of conduct. But the mere fact that our aspirations can bias our actions, as well as our assessment of such actions, is an important mechanism to ensure self-reflective forms of action control (Paglieri and Castelfranchi 2008; Castelfranchi, 2012). And Castelfranchi has argued (1995), with good reason, that such higher mechanisms for behavioral regulation enables human beings to achieve feats that would be impossible for simpler (and thus less biased) creatures.

[8] Genuine instances of defensive beliefs should be kept apart from other non-evidential constraints over belief formation and change, such as ensuring minimal change (Alchourrón, Gärdenfors and Makinson 1985; Harman 1986). The parsimonious tendency to minimize the amount of cognitive restructuring required to accommodate new information certainly exerts a biasing influence over belief dynamics, and this might on occasion have pernicious effects, even if it is highly convenient on average. But such bias, and others like it, depends on an imperfect attempt to optimize efficacy and is not born out of fear: thus it does not relate to scary doubts at all.

The pragmatic defense of motivationally biased beliefs presented by Castelfranchi and colleagues is not new, albeit its details are. The basic rationale of it is similar to that of William James' much celebrated essay on "The will to believe" (1895/1979), and indeed this line of reasoning is deeply rooted in and inspired by pragmatism. In short, undertaking the best possible course of action is seen as rationally superior to believing the truth, at least when these two values come into conflict, as exemplified by Pascal's famous wager. The standard counterargument against this defense of defensive beliefs (pun intended) is that the pragmatic efficacy of such biases do not make them any more evidence-based or truth-preserving, hence it does not constitute a valid form of epistemological justification. A mistake of reasoning remains a mistake of reasoning, even if it happens to save your life.

A way out of this objection is to make the legitimacy of motivational biases conditional on the lack of better epistemological alternatives: when you do not have ways to solve your doubt on the grounds of evidence, it is rationally legitimate to let your goals shape your beliefs. This is the position James himself endorses, when he writes:

> Our passional nature not only lawfully may, but must, decide an option between propositions, whenever it is a genuine option that cannot by its nature be decided on intellectual grounds; for to say, under such circumstances, "Do not decide, but leave the question open", is itself a passional decision – just like deciding yes or no –, and is attended with the same risk of losing the truth (James 1895/1979, p. 11).

In addition to that, James also suggests that instances where belief can be reached on purely intellectual grounds (i.e. based solely on evidence) are much rarer than what is usually assumed, even in science,[9] so that most of our beliefs must in fact rely, to some extent, on our volitional dispositions – that is, on our goals and values.[10] Whether or not one agrees with James regarding the relative infrequency of pure evidence-based belief, it is still possible to endorse his circumscribed defense of motivation-biased belief: insofar as evidence do not suggest either belief or disbelief, there is nothing epistemologically wrong in letting our goals settle our convictions. To which a stalwart critic might object that lack of epistemological fault does not imply presence of epistemological merit: granted, there is nothing wrong in letting your goals bias your beliefs in the absence of evidence, but this is not to say that there is something *right* in that process – apart from whatever pragmatic benefits you might achieve, which is beside the point for our imaginary critic. In this view, proper justification requires that a belief forming procedure has some epistemic virtue, whereas

[9] Although it does not bear on the present discussion, it is worth recalling James' penetrating caricature of scientific method, as presented in his essay, since it is still actual nowadays, possibly more than ever: «The most useful investigator, because the most sensitive observer, is always he whose eager interest in one side of the question is balanced by an equally keen nervousness lest he become deceived. Science has organized this nervousness into a regular technique, her so-called method of verification; and she has fallen so deeply in love with the method that one may even say she has ceased to care for truth by itself at all. It is only truth as technically verified that interests her. The truth of truths might come in merely affirmative form, and she would decline to touch it».

[10] For in depth analysis of the complex relationship between goals and values, see Miceli and Castelfranchi (1989).

lack of obvious epistemic vice is not a sufficient condition. Hence, the will to believe would not qualify as proper justification, not even in the absence of evidence, contra James.

This debate, far from being settled, concerns the normative validity (or lack thereof) of cognitive biases, in this case the role of scary doubts in prompting the formation of defensive beliefs: as such, it pertains more to philosophical epistemology than to psychology. However, also psychologists could (and often do, see for instance Gigerenzer and colleagues) wonder about the merits and shortcomings of such biases. For them, the debate is more naturally framed in terms of efficacy and adaptation.[11] To what extent defensive beliefs happen to be correct and thus successful in supporting one's behavior? And is it possible to track the evolutionary processes that might have produced these belief-formation practices as an adaptation to certain selective pressures? In order to address these questions, it is not enough to establish that, in general, goals do exert a systematic influence over our beliefs: we also need to specify more clearly what kind of influence is exerted, and what mechanisms are responsible for it.

Without entering in too much details here, it is fair to say that the work of Castelfranchi and collaborators provides some guidance in this respect. In a series of papers devoted to analyze belief dynamics and their interaction with goals (Castelfranchi 1996; 1997; 2004; Paglieri 2004; 2005; 2006; Paglieri and Castelfranchi 2005; 2006; 2007), they indicate two distinct ways in which goals might affect belief:[12] *relevance* and *likeability*. These are conceived as properties of information, that is, of data that the agent is aware of without yet being committed to. Both the relevance and the likeability of an information are dependent upon the agent's goals and might influence the likelihood of believing that information, but they do so in very different ways. More exactly, relevance is defined as the *pragmatic utility* of an information, i.e. the number and value of the (pursued) goals that depend on reaching a conclusion, i.e. a belief, on that matter; in contrast, likeability is understood as the *motivational appeal* of an information, that is, the value of the pursued goal(s) directly fulfilled by believing it. Likeability arises because, subjectively speaking, a goal p is satisfied only when the agent believes p, whether or not such belief is factually correct – in short, goals are satisfied by beliefs, not by states of the world. This puts a definite pressure on believing certain things to be true, insofar as doing so would directly and immediately satisfy the agent's current goals. Relevance, on the other hand, works very differently: it directs our attention towards certain information rather than others, depending on current needs, but without pressuring us to believe them. This still has a definite influence on belief formation, since it determines what contents come to be

[11] It is worth emphasizing that there is an intriguing, yet largely unexplored continuity between the pragmatist focus on practical success and the evolutionary concern with the adaptive value of certain behaviors or traits. Even though these are different notions used in different debates within different disciplines, they share the same core: the idea that success (pragmatic in one case, evolutionary in the other) is the ultimate end, whereas truth is just a means for it.

[12] For the converse path of influence, i.e. from beliefs to goals, see Castelfranchi and Paglieri (2007).

considered as candidates for belief, but does not make us any less prone to wishful thinking (for further discussion on this distinction, see Paglieri 2005).

The distinction between relevance and likeability provides us with a simple framework to analyze the different impact that Peircean doubts and scary doubts have on belief formation, and also to question separately the adaptive value of these different mechanisms of goal-directed belief formation. It is easy to see that Peircean doubts, and our reaction to them, concerns relevance alone: in fact, relevance is precisely a measure of how much a certain information is needed to advance the agent's current plans, that is, of how many Peircean doubts would be quenched by the corresponding belief, and how important they are for the agent. It is also apparent that this bias is adaptively beneficial, since beliefs are meant to serve as "maps by which we steer" (Ramsey 1931), and the general direction of our behavior is obviously determined by goals. So it is certainly a good thing that relevance prompt us to focus on goal-related matters, thus keeping goals and beliefs in sync. Were it not so, we would find ourselves ill equipped to pursue our current interests: for instance, we might end up walking down the aisle at our own wedding with the head full of thoughts on differential equations, but without a clue on how to behave with our spouse-to-be. Indeed, when similar incidents happen, they are universally regarded as disastrous, and diagnosed as bad cases of epistemic distraction.[13] I suggest that their root cause is a failure of relevance.

Likeability, on the other hand, is uniquely tied to scary doubts and defensive beliefs: it is precisely when an information happens to be likeable or distasteful for us (i.e. prone to satisfy or frustrate our goals) that we feel the pressure to either believe or disbelieve it, independently from whatever evidence we might have on the matter. Whether or not this tendency is adaptive constitutes a thorny issue, that I would spell out as follows: Assuming likeability-induced belief formation works as supposed, is it an effective mechanism, and in which sense? It is certainly not infallible, since it can easily lead us to entertain false beliefs (e.g. an overinflated opinion of ourselves, or a mistaken faith in the love of our relatives). But is it typically fallacious, that is, does it produce more false beliefs than true ones? This is an empirical question, one which will be very difficult to answer and far exceeds the purpose of the present contribution.[14]

By way of mere speculation, I just want to stress that from the (evident) fact that likeability often biases belief formation we should *not* immediately infer that such process must have some evolutionary relevance or adaptive value.

[13] By epistemic distraction I mean here an instance where the agent has a unique or dominant goal, but fails to focus on the beliefs needed to pursue it effectively. Another type of distraction can be labeled as motivational, and refers to instances where more than one goal compete to catch the agent's attention, so s/he ends up switching back and forth between different tasks – cyber-slacking being a clear example of this phenomenon.

[14] According to Mercier and Sperber (2011), this question is actually besides the point, because the function of reasoning is not to deliver true belief, but to make us capable of producing good arguments and assessing poor ones. This would make our belief system highly adaptive, even if it turned out to produce a substantial body of falsehoods – that would not been its purpose, after all. I am somewhat skeptical of this radical proposal, but I acknowledge that it provides an interesting alternative to the standard account of belief, worthy of greater attention in future work.

This is of course true in general when we reason in evolutionary terms (not all current traits are adaptations, for sure), but there are also special reasons to be extra careful in assessing scary doubts and defensive beliefs. This is why: as mentioned, the likeability bias is a side product of a basic feature of goal-directed behavioral control, namely, the fact that *goals are subjectively satisfied by beliefs*, not by facts. This is unavoidable, insofar as our perception of external reality is filtered by our representational system. But it also creates a vulnerability, because the belief that originally serves as a mere means to check that the world has been changed according to plan, can later become an end in itself, since goal satisfaction is in fact dependent on it. This is an instance of a more general phenomenon in the dynamics of signs: once a stable association is determined between a sign and its referent, so that we start reacting to the sign as we would to the referent itself, then producing the sign is sufficient to elicit our behavioral response. When we run out of the building to save our life, we are reacting to the alarm, not to the fire – even if of course the reason for doing so is because we assume the fire is there. Beliefs suffer the same fate: insofar as they are conceived as reliable signs of the attainment or failure of our goals (and we cannot help but conceiving them so), then the danger of likeability bias immediately lurks in their shadow.

If this is correct, then the quest for an evolutionary justification of scary doubts and defensive beliefs might well be a fool's errand. Regardless of how often such beliefs end up being correct or incorrect, the mechanism that produces them would be more parsimoniously understood as a side effect of a much larger adaptation – namely, our goal-oriented, cognitively mediated action control system (Pezzulo and Castelfranchi 2009; Pezzulo this volume).

4 No doubts: the mind as a coherence-seeking device

Phenomenologically speaking, there is a sharp difference between a state of doubt and a condition of ignorance. When in doubt, we are typically presented with a finite number of competing options, all of which appear to us equally plausible – or, at least, close enough in plausibility as to preclude a principled choice in favor of one of them. Ignorance feels different: when we ignore something, we are mostly in the dark about it, so that no option at all presents itself as a candidate for belief or action. In spite of this phenomenological difference, doubt and ignorance share something important: they both carries a *potential for internal inconsistency*, that is, for the possibility of holding two or more conflicting mental states. Such inconsistency is more manifest in our doubts and more hidden in our ignorance, but the potential for it is present in both cases. The ignorant is by definition open to multiple possibilities (typically far more than those entertained by a doubtful mind), and most of these options are known to be mutually incompatible. Hence both doubt and ignorance can be the harbinger of inconsistency.

This is relevant for our cognitive practices, not because our mind is or should be free of contradictions, but because for some classes of mental states, and only for them, we have a natural tendency to strive for coherence by removing inconsistencies.[15] On this view (Castelfranchi 2004; Paglieri 2006), the mind is conceived as a *coherence-seeking device*, which is different from a coherence-bounded machine: in particular, the quest for coherence regards both epistemic and motivational attitudes, but is also limited to those attitudes that entail some sort of commitment for the agent, i.e. beliefs and intentions, respectively; with regard to more basic attitudes, such as data and desires, contradictions do not pose any problem for our mind. Figure 1 summarizes this parallel between the dynamics of epistemic and motivational states, whereas the details of each process has been discussed in details elsewhere (see Paglieri 2004 for an account of data-oriented belief formation, and Castelfranchi and Paglieri 2007 for an analysis of belief-based goal processing).

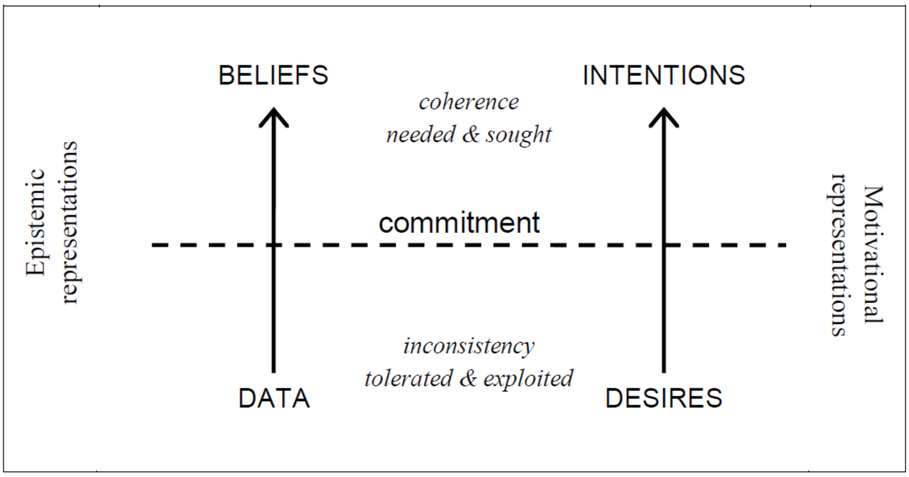

Fig. 1 The mind as a coherence-seeking device

It is the need to act effectively upon reality that prompts us to commit ourselves to certain intentions and to endorse certain beliefs, moving beyond ignorance and doubt – sometimes by an act of will, but mostly as a reaction to whatever evidence we are presented with. It is in this sense that Peircean doubts work as an irritant for cognition: they simultaneously indicate that something needs to be ascertained in order to act, and also that there is the potential for an inconsistency in our belief system, unless the matter is settled.

[15] This is typically achieved either by making a choice among multiple conflicting candidates, or by compartmentalizing, that is, segregating conflicting mental states into rigidly separated areas of our mind (see Cherniak 1986 for in-depth discussion of the role of compartmentalization in human rationality).

5 Conclusions

In this paper I took advantage of Peirce's apt metaphor to distinguish two types of doubt, and to analyze their different impact on belief formation. In particular, I argued that Peircean doubts work as a beneficial irritant to instigate informed guesses that are necessary for effective action, even though they lead us occasionally astray. In contrast, scary doubts pressure us towards self-serving beliefs, often regardless of factual evidence: this might or might not be warranted on practical or prudential grounds, but it is certainly risky business as a general epistemic habit. However, it is a vulnerability in our belief formation system that we have to live with, insofar as it originates from the practical need to (fallibly) determine whether a certain goal is attained. This implies that "scratching our beliefs" (that is, letting our goals influence our convictions) is always reasonable in the face of Peircean doubts, while it might be a ruinous path down a slippery slope in matters of scary doubts. Unfortunately, Peirce's metaphor turns out to be devilishly accurate: as for most irritations, refraining from scratching our beliefs in the presence of doubts is extremely hard, regardless of what made us itchy.

In unfolding this view of the connection between doubts, beliefs, and actions, I also tried to weave it within the rich fabric of Castelfranchi's theory of cognition and, more precisely, within his analysis of belief formation and goal processing. The purpose of this exercise was not only to honor Castelfranchi's contribution to this area of inquiry, as it is customary on such an occasion, but also to unearth a conceptual root that Peirce's pragmatism and Castelfranchi's goal theory have in common. This is the *primacy of action* in our understanding of the mind. For both these scholars, possibly for different reasons, cognition is *for* action, first and foremost. It is so not only because goals are mental attitudes on a par with beliefs (a lesson Castelfranchi never tires of hammering into the mind of students and colleagues), but also because belief formation itself is motivated and shaped by goals. Peirce's irritation of doubt, far from being the intellectual, methodological doubt advocated by Descartes, is precisely a manifestation of how goals initiate and guide all our cognitive undertakings, including the formation and revision of beliefs. Thus I dare to say that Peirce and Castelfranchi would have shared the same doubts on any theory of cognition that takes truth, rather than success, as its cornerstone.

References

1. Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic* 50, pp. 510-530.
2. Alicke, M., & Sedikides, C. (2011). *Handbook of self-enhancement and self-protection*. New York: Guilford Press.
3. Aronson, E. (1969). The theory of cognitive dissonance: a current perspective. In L. Berkowitz (Ed.), *Advances in experimental social psychology, vol. 4* (pp. 1-34). New York: Academic Press.

4. Bem, D. (1967). Self-perception: an alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*, 183-200.
5. Bratman, M. (1992). Practical reasoning and acceptance in a context. *Mind, 101*, 1-15.
6. Castelfranchi, C. (1995). Self-awareness: notes for a computational theory of intrapsychic social interaction. In G. Trautteur (Ed.), *Consciousness* (pp. 55-80). Napoli: Bibliopolis.
7. Castelfranchi, C. (2004). Reasons to believe: cognitive models of belief change. Manuscript. Ws *Changing minds: cognitive, computational and logical approaches to belief change*, Amsterdam 29/10/2004. `http://www.unisi.it/ricerca/dip/fil_sc_soc/dot-sc/castelfranchi.pdf`. Accessed 8 June 2011.
8. Castelfranchi, C. (2012). "My mind": reflexive sociality and its cognitive tools. In F. Paglieri (Ed.), *Consciousness in interaction: the role of the natural and social context in shaping consciousness* (pp. 125–149). Amsterdam: John Benjamins.
9. Castelfranchi, C., & Paglieri, F. (2007). The role of beliefs in goal dynamics: prolegomena to a constructive theory of intentions. *Synthese, 155*, 237-263.
10. Castelfranchi, C. (1996). Reasons: belief support and goal dynamics. *Mathware & Soft Computing, 3*, 233-247.
11. Castelfranchi, C. (1997). Representation and integration of multiple knowledge sources: issues and questions. In V. Cantoni, V. Di Gesù, A. Setti & D. Tegolo (Eds.), *Human & machine perception: information fusion* (pp. 235-254). New York: Plenum Press.
12. Cherniak, C. (1986). *Minimal rationality*. Cambridge: MIT Press.
13. Cohen, L. (1989). Belief and acceptance. *Mind, 98*, 367-389.
14. Dempster, A. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics, 38*, 325-339.
15. Elster, J. (1983). *Sour grapes: studies in the subversion of rationality*. Cambridge: Cambridge University Press.
16. Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
17. Gigerenzer, G., Todd, P., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
18. Greenwald, A. (1980). The totalitarian ego: fabrication and revision of personal history. *American Psychologist, 35*, 603-618.
19. Harman, G. (1986). *Changes in view: principles of reasoning*. Cambridge: MIT Press.
20. James, W. (1895/1979). *The will to believe and other essays in popular philosophy*. Cambridge: Harvard University Press.
21. Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.
22. Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480-498.
23. Magnani, L. (2009). *Abductive cognition. The epistemological and eco-cognitive dimensions of hypothetical reasoning*. Berlin: Springer.
24. Magnani, L., & Bertolotti, T. (2011). Cognitive bubbles and firewalls: epistemic immunizations in human reasoning. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of CogSci 2011*. Accepted poster, forthcoming.
25. Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34 (2)*, 57-74.
26. Miceli, M., & Castelfranchi, C. (1989). A cognitive approach to values. *Journal for the Theory of Social Behaviour, 19*, 169-193.
27. Miceli, M., & Castelfranchi, C. (2012). Coherence of conduct and the self-image. In F. Paglieri (Ed.), *Consciousness in interaction: the role of the natural and social context in shaping consciousness* (pp. 151–177). Amsterdam: John Benjamins.
28. Paglieri, F. (2009). Acceptance as conditional disposition. In A. Hieke & H. Leitgeb (Eds.), *Reduction: between the mind and the brain* (pp. 29-49). Berlin: Ontos-Verlag.
29. Paglieri, F., & Castelfranchi, C. (2008). Cambiare la mente: mindreading, azione intenzionale e coscienza. *Sistemi Intelligenti, 20 (3)*, 489-520.
30. Paglieri, F. (2004). Data-oriented belief revision: towards a unified theory of epistemic processing. In E. Onaindia & S. Staab (Eds.), *Proceedings of STAIRS 2004* (pp. 179-190). Amsterdam: IOS Press.

31. Paglieri, F. (2005). See what you want, believe what you like: relevance and likeability in belief dynamics. In L. Cañamero (Ed.), *Agents that want and like: motivational and emotional roots of cognition and action* (pp. 90-97). Hatfield: AISB.
32. Paglieri, F. (2006). *Belief dynamics: from formal models to cognitive architectures, and back again*. PhD dissertation, University of Siena. `http://www.media.unisi.it/cirg/fp/Paglieri_phd_thesis.pdf`. Accessed 8 June 2011.
33. Paglieri, F., & Castelfranchi, C. (2005). Revising beliefs through arguments: bridging the gap between belief revision and argumentation in MAS. In placeI. Rahwan, P. Moratïs & C. Reed (Eds.), *Argumentation in multi-agent systems* (pp. 78-94). Berlin: Springer-Verlag.
34. Paglieri, F., & Castelfranchi, C. (2006). The Toulmin test: framing argumentation within belief revision theories. In D. Hitchcock & B. Verheij (Eds.), *Arguing on the Toulmin model* (pp. 359-377). Berlin: Springer.
35. Paglieri, F., & Castelfranchi, C. (2007). Belief and acceptance in argumentation. Towards an epistemological taxonomy of the uses of argument. In J. A. Blair, F. H. van Eemeren & C. A. Willard (Eds.), *Proceedings of ISSA 2006* (pp. 1011-1018). Amsterdam: Sic Sat.
36. Peirce, C. S. (1958). *Charles S. Peirce: selected writings*. New York: Dover.
37. Pezzulo, G., & Castelfranchi, C. (2009). Thinking as the control of imagination: a conceptual framework for goal-directed systems. *Psychological Research, 73*, 559-577.
38. Pezzulo, G. (in press). Re-founding cognitivism based on the cybernetic idea of goal-directed action. This volume.
39. Ramsey, F. (1931). Truth and probability. Reprinted in H. E. Kyburg & H. E. Smokler (Eds.) (1964), *Studies in subjective probability* (pp. 61-92). New York: Wiley.
40. Shafer, G. (1976). *A mathematical theory of evidence*. Princeton: Princeton University Press.
41. Swann, W. Jr., & Read, S. (1981). Self-verification processes: how we sustain our self-conceptions. *Journal of Experimental Social Psychology, 17*, 351-372.
42. Swann, W. Jr. (1999). *Resilient identities: self, relationships, and the construction of social reality*. Basic Books: New York.
43. Taylor, S., & Brown, J. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin, 103 (2)*, 193-210.
44. Thagard, P. (2004). What is doubt and when is it reasonable? In M. Ezcurdia, R. Stainton & C. Viger (Eds.), *New essays in the philosophy of language and mind* (pp. 391-406). Calgary: University of Calgary Press.
45. van Fraassen, B. (2001). Constructive empiricism now. *Philosophical Studies, 106*, 151-170.
46. Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12 (3)*, 129-140,
47. Woods, J. (2005). Epistemic bubbles. In S. Artemov, H. Barringer, A. Garcez, L. Lamb & J. Woods (Eds.), *We will show them: essays in honour of Dov Gabbay, vol. 2* (pp. 731-774). London: College Publications.