Commonsense Reasoning

Do Humans Think?

Jo Bervoets, 2005.
Master's Thesis for "Master in Artificial Intelligence", Option "Cognitive Sciences".
Katholieke Universiteit Leuven
Correspondence to: jo.bervoets@skynet.be

To my wife.
I gave up these pursuits for her
and got back
endlessly more, times four.
Then she did not stir
when I went back,
for a little while,
to this, my initial, track.

“Chapeau!” my dear Els.

I. Table of Contents

II. Preface

From when I was very young I thought long and hard about opposites. There was something in it that continued to baffle me. That something still doesn't cease to amaze me. Opposites are so central and so simple. From our very first days in school we are taught how pure the concept of opposites is. Yet there is not a single pair of opposites that admits of a clean application to reality. As we all know the extreme right and the extreme left share a great number of features. They seem to share more features with each other than they do with the supposedly neutral point in the centre.

In my adolescent mind there consequently grew a deep suspicion of the purity of logic. Somehow, like with opposites and excluded middles, the laws of logic always hold but never apply. Behind all corners of an argument there seemed to lurk an unknown quality that could not be tamed by logic. I radically refused any extremism that, in one or another way, seemed to come from a mistaken desire to regiment nature according to the unnaturally strict rules of logic. I read fiction, Kierkegaard and Nietzsche.

Afterwards, it seemed to me that what appeared to be a complete opposite of regimentation was equally unattractive. Mysticism and belief in magically transcendent qualities defying explanation led into the at least as unattractive option of exclusion based on coincidental features. Ironically, the more elements of a scientific nature are included in such mystic belief systems the more awful the arbitrariness of such exclusion becomes. I read non-fiction, Wilde and Montaigne.

It is in these opposites that the source of all irony lays. Only in laughter can we deal humanely with this condition of being eternally suspended between the pure and the practical. What is needed, or at least so I thought, was a concrete understanding of humans as such instead of an abstract treatment of how or what humanity should be. This brought me to the cognitive sciences, only to find in the latter again the same types of opposites, now in miniature. The same type of crossover confusions as well and also, unfortunately, much support for both regimenters and arbiters.

The source of production and creation, according to some, lies in recursion and in computation. Others, always glancing to the cookbook of the computationalists, rather find some irreducible quality needs to be responsible for this magical thing humans are. The first tradition leads some to prefer a positive manifold related to the purer mind. As it happens, it is the type of mind most akin to theirs. The second keeps its eyes, no doubt, firmly focused on what's neither here nor there. As of yet, luckily for us, no revelations have, as of yet, fixed the attention on an arbitrary here or there.

So, there is no escaping it: I need to face both my Scylla ánd my Charibdis. Neither regimentation, nor arbitrariness can have their way. There can again be no exclusion of the middle. It is laughter and irony that are productive and creative, not recursion. It is the middle that is of interest to us. It is the middle I will explore in this thesis.

As in other journeys where one wants to explore something new and hence only knows on beforehand what one is to avoid, I also came to navigate in a way opposite to the way I started out with. Hindsight has not cancelled this out completely from the text. I hope there is more excitement than annoyance in seeing facts determine the journey.

Ladies and Gentlemen, this is my soapbox! But I do not apologize, not even to my forced readership, for much of the poetic license contained in this volume. I only apologize for not having more years to clean out all the corners. I hope not too many will stumble and fall over the dirt and chaos remaining, here and there, on the floor. I would say: do as if you were at home and skip, if necessary, through to the bits of text of specific interest mostly in the middle of the volume.

My sincerest thanks to Kristien Dieussaert for helping me look in what I have come to consider as the right direction.

III. Abstract

"Common sense is in no way capable of explaining the world"

Bertrand Russell

And neither is common sense capable of explaining commonsense reasoning. The purpose of this thesis can most briefly be stated as follows: commonsense reasoning is an oxymoron. Common sense doesn't reduce to the regimented computations underlying formal logico-mathematical systems. It is not at all a type of reasoning in this, the preferred, sense of the word. Neither is it a superior short cut allowing to "see" more directly the outcome of computations without needing to make them. Russell was far more right about all that then history has conceded up to now.

I will argue that humans do not compute in any meaningful sense of that term. It is often held that such computations are fundamental to reasoning and thinking. If so, we need to answer the basic question - "Can machines think?" - affirmatively. But, by the same token, we need to answer the mirror question - "Do humans think?"– Negatively. The basic conjecture in this work is: thinking as expressed in logico-mathematical formalisms is not psychologically plausible – neither literally nor as an ideal we tend to approximate. Rather, reason is something we have produced collectively through the social criticism characteristic of scientific enterprise. The truth about common sense lies elsewhere.

We should not let ourselves become confused by the strength of common sense to absorb the outcomes of socially constructed formal mechanisms. The immediate appeal of the functionalist claims about computations over representations of the world is like that of a fata morgana. Setting us as some kind of Dr. Spock is "the Vulcan fallacy of cognitive psychology". Exploring the antinomies to which such a fallacy leads us is the scope of the first chapter.

The purpose of this thesis more elaborately stated is to find a formal treatment of where one can look for common sense, properly understood, and how this phenomenon of "practical reason" allows us socially to make contact with the type of "pure reason" allowing any critical enterprise. I will do so by using relatively recent work of pure reason, by D. Edgington, H. Kyburg Jr. and H. P. Grice. The metaphor used is that of building a plan for "the house of reason".

Common sense *per se* consists in navigating various rooms jam-packed with background knowledge. The rooms are built from conditional statements to be interpreted, following D. Edgington, as Bayesian conditional beliefs (Chapter 2). The material on which the house is built is unlike the material on which logic is built. Common sense needs to be built on uncertainty and the foundation of our house therefore needs to be the principled way of performing uncertain inferences proposed by H. Kyburg Jr. (Chapter 3). Finally our house needs a roof that fits the rest of the construction but that also allows us definite, certain, shelter. This roof is made from the conventional meanings that are socially constructed in ways suggested by an often-neglected link between conventional and non-conventional meanings quesserted by H. P. Grice (Chapter 4).

The validity of each of these sub-conjectures is explored in the indicated chapters with reference to the established body of empirical evidence in the broad field of psychological reasoning literature. A more comprehensive test of the over-all conjecture is attempted in Chapter 5. Both a suggestion of a possible empirical dissociation between conventional (computational) and non-conventional (commonsensical) reasoning is given. The attempt is admittedly embryonic at best.

Finally, in Chapter 6, we consider something quite odd about the house we have built. It is not the type of house standing statically as a supreme but unattainable Truth. Quite the contrary, this house is very dynamic, expanding (or, alas, contracting) through the interaction of mutual criticism and individual creativity. Conventional logic used in social interaction and creative power of an individual's common sense cannot be reduced to each other. One does justice to neither when focusing only on the one side of recursive compositionality or on the other side of intuitive creativity. Individuality derives from sociality just like any advanced social behavior derives from individuality. Although only briefly sketched here, I hope to convey to the reader that the moral of this work is moral.

Chapter 1 – Towers of reason vs. reasonable housing:

"We have found, indeed, that although we had contemplated building a tower which should reach to the heavens, the supply of materials suffices only for a dwelling-house, just sufficiently commodious for our business on the level of experience, and just sufficiently high to allow of our overlooking it. The bold undertaking that we had designed is thus bound to fail through lack of material -- not to mention the babel of tongues, which inevitably gives rise to disputes among the workers in regard to the plan to be followed, and which must end by scattering them over all the world, leaving each to erect a separate building for himself, according to his own design. At present, however, we are concerned not so much with the materials as with the plan; and inasmuch as we have been warned not to venture at random upon a blind project which may be altogether beyond our capacities, and yet cannot well abstain from building a secure home for ourselves, we must plan our building in conformity with the material which is given to us, and which is also at the same time appropriate to our needs. "

Immanuel Kant, Critique of Pure Reason, 1966.

The debates within the cognitive sciences pretty much feel the same way, as the Scholastic debates must have felt to I. Kant. Divisions inspired by strong principles – see Stanovich & West (2000), Samuels et al. (1999), inter alias – alongside a substantial fragmentation on most specific matters. Specifically with reference to so called "higher" cognition, or reasoning, one finds rather heated debates between a more rationalist point of view exemplified by Stanovich & West (2000) and a more empiricist view as, for instance, adopted by Gigerenzer & Selten (1999).

My, maybe somewhat overambitious, contention is that by and large the analogy holds. The divisions and debates referred to above mask a more profound issue similar to the one Kant dealt with. There is a common source to the issues encountered by either side in these debates. This common source consists in the shared view that the mind should be able to mirror, in principle, the external human behavior of reasoning. In other words, the brain is a miniature computer of sorts on which all overt reasoning can be implemented and the brain is, as well, the source of the mind. The most direct example of an argument to this effect can maybe be seen in Barsalou (1999) arguing from perceptual symbols to reasoning with abstract propositions.

In the terms of the quote above, I hold that these views are "bound to fail by lack of material", in the brain. There is more to the social practice of reasoning than can be accounted for by the mere operation of the brain. At the same time there is more to the operation of the brain than can be functionally accounted for using the basic terminology of pure reason.

I highlight in a first section how traditional accounts, apparently opposite to each other, each produce unsolvable antinomies – grounded in foundational terminology such as representations, computations and connections – allowing no account of the key concept of common sense. I hope this section is read with some charity, it is at this point not much more than a first attempt. In the second section, I will introduce the basic outline of an alternative account whose ambition it is to be more modest insofar as the human cerebral capacities are concerned. This account centers on explaining the phenomenon of common sense as the primary and central "thing-to-be-explained". I hope this new plan-of-the-house to be plausible enough to account for how a human can start from basic perceptual experience to end up in writing a thesis explaining how humans can start from basic perceptual experience to end up writing my thesis. But that will be argued substantively only in the later chapters.

1.1. The antinomies of towering reasoning:

The similarity of these antinomies to the Kantian antinomies is currently limited to the fact that both of them number four. I nevertheless use the terminology to show the type of argument I think is needed in deconstructing the traditional views on human reasoning.

1.1.1. Common sense: The Fodorian antinomy

The computational/modular approach leads to an insurmountable problem as eloquently admitted by J. Fodor, one of the founding fathers of the approach. In Fodor (2001), he shows the "frame problem" to be an intractable problem on the computational/modular hypothesis, preferred by many. Common sense emerges as an almost magical fact of life, because it fairs well with computationally "hard" problems and defies being neatly analyzed in modular terms. A similar analysis has often been made in reasoning literature when considering "complete" problems for which solutions cannot be computed in finite time – see, for instance, Oaksford & Chater (1993). In fact, this issue is foundational from Newell on in the functionalist tradition and has led to the "heuristics/biases" and "bounded rationality" traditions (as can be seen from Evans (1993), Kahneman (2003)).

Common sense then simply cannot be explained by symbol-manipulating devices inside one's head, let alone by an "organ" for first-order predicate logic as is often the implicit assumption made[1]. E. J. Lowe (1993) makes essentially the same point from a more philosophical angle: symbolic reasoning, whether according to formal syntactics or semantics, cannot account for human reasoning. The "language of thought" hypothesis – and with it mentalistic or cognitivist approaches – seem to be in danger from the inside out. Nonmonotonic logics (see treatment in e.g. H. Rott (2001)) maybe can be devised to fit the phenomena but only at the expense of increasing psychological implausibility.

Consequently, many may be tempted to conclude with Lowe that we need to look into "the direction of a 'connectionist' response". This in order to come to grips with "the 'holistic nature' of propositional understanding" (Oaksford & Chater (1993)). However, it is difficult to imagine how holism could be a solution. The frame problem as presented in Fodor (2001) does not seem to consist in an ability to look at the whole of the data. Nor does it seem that adopting "other kinds of" processing solutions provides us with the necessary qualitative difference of what is essentially a "hard" mathematical fact. Finally, it seems to lead to a quite odd inversion with respect to most intuitions on cognitive architecture – as, for instance, per Kahneman (2003) – to equate "higher" cognition with a non-analytic core whilst allowing with Fodor the periphery to be based on computationality.

The core issue is that through common sense we are able to look directly at "appropriate" subsets of the total potentially relevant knowledge. We seem to need a rather less than more holistic solution if we're to make sense of common sense. Take, for instance, the well-known relevance account of Sperber & Wilson (1986), the intuitive appeal of this account largely consists in its attempt at explaining how one can rationally restrict the frame of reference in any given context. None of the traditional views, see also Fodor (2001) on Sperber & Wilson (1986), succeed however in accounting for such relevance

The computational account seems to lead to the need for positing a holistic non-computational core of cognition whereas the connectionist approach is at pains to account for an analytic core of cognition. In a boutade one might conclude that AI should really be called OI – "Other Intelligence" – stressing that it not only solved problems different from the ones solved by humans but that it also does so in a quite different manner.

1.1.2. Intuitions: the Procrustean antinomy

Following the neo-rationalist approach there is a straightforward "measure" of rationality: the canon of logic (if one can use the definite article here). One can then think of "measuring" rationality against the standards of reasoning as given by that canon – something that is taken almost literally in Stanovich & West (2000). Human intuition then is by assumption suspect and the yardstick to measure rationality by the same token unproblematic. As Russell probably would have had it, human rationality should get the treatment of being fitted in the Procrustean bed of a sound and complete logic.

This maybe explains why much of the reasoning literature of the past couple of decades would more appropriately be called the "irrationality" instead of "heuristics-biases" tradition. Nevertheless, as for instance pointed out by Evans (1993), it may be unfair to demand this measure of rationality to be used for human reasoning. Rather, one may have cause to consider the human case as the proper benchmark

[1] As much should have been clear after the many attempts to make a "purely" declarative programming language resulting in textbooks stressing the importance of programmers not forgetting the procedural "meaning" of their programmes, see Bratko (2001).

as it is obviously more performant within what evolutionary psychologists call "the proper" or "the ecologically relevant" domain in which it needs to excel – see Gigerenzer in Gigerenzer (2002).

Nevertheless, the debate between those believing intuitions are benevolent Muses and those believing intuitions are somehow suspect is more like the discussion between members of the same party than as the discussion between two real opposites. Indeed, the foundational term in this debate is "bounded rationality" (a term coined by the common ancestor, Newell, of both sides of the debate). The fact that Gigerenzer and some evolutionary psychologists stress the fact that attaining the computational norm isn't ecologically sensible does not change the fact that the norm is still relied on as the ultimate benchmark. Their use of "heuristic" appears to be interchangeable with the others use of "fallacy".

In fact, without some fixed normative standard as a benchmark there is not much sense in the word "heuristic". One risks, with a lot of post hoc explanations, to fit the norm to the Procrustean bed of the unlimited amount of specific adaptations to specific environments (not dissimilar to modern practices in formal semantics of natural language – see Chierchia (1996)). Based on the unproblematic adoption of classical logic one may well be led to associate intelligence with repetition of socially accepted ways of (linear) reasoning (as in the case of 11-year old children believing 4 eggs need to boil 4 times longer than a single egg). But based on the unproblematic adoption of "adapted" behavior as best behavior one may well end up associating intelligence with mere survival[2].

In the meantime, few normative theories are immune to change (see Chapter 2 for a case in point), and much of the revisions are ultimately justified by appeals to human intuition as the final arbiter – a type of intuitions that nobody with common sense would be able to deny. Hence, one finds in recent work by H. Rott (2001) a prototypical philosophical intuitive defense of such and other postulates whilst on the other hand psychologists (Ford & Billington (2000) and, for a full review, Dieussaert (2002)) try to measure how intuitive such postulates actually are. In other words, one risks getting bogged down into a situation of "*Your intuition against mine*"*, or "Your logic against mine"*. A risk that is exacerbated if some people's intuitions are put up as more trustworthy than those of others as, for instance, done in Benferhat et al. (2002) and strongly suggested in Stanovich & West (2000) (at least if one wants to use evaluative concepts like "rational" and, after all, it would be counterintuitive not to – see Chapter 3).

Starting out from unproblematic normativity one seems forced to label as irrational the same intuitive behavior that seems both the most excellent and most difficult to artificially recreate, whilst starting from unproblematic intuitions leads us into an unsustainable absence of normative rationality. The head and its brain seem too small a place to sustain both intuitions and normativity (whether bounded or not) as unproblematic and primary.

1.1.3. "Rational or animal?": the Cartesian antinomy

Once the definition of man as a rational animal is given, explanations tend to bifurcate into explaining either element of the definition separately. Whilst Cartesian dualism is no longer over(t)ly popular, a range of quasi-dualist hypothetical cognitive architectures remains prominent (the S1/S2 systems as per Stanovich & West (2000) and Kahneman (2003) and a similar architecture as per Donovan & Epstein (1997) referring to the "analytic and intuitive" modes of reasoning). As in classical dualism, they come with the central problem of interfacing between two systems that are almost ontologically different in assuming a hard dichotomy between, roughly, cognition and perception.

From the anim(al)istic side empiricist research applies the biological concepts of ecological adaptation almost anywhere. A characteristic proposal in this vein is the one on perceptual symbols in Barsalou (1999), which attempts to reconstruct "bottom-up" the whole machinery of classical propositional logic out of percepts. As I argued elsewhere (Jo Bervoets (2005a), see also OPC commentary of A. J. Wells on Barsalou (1999)), the Barsalovian account overstretches the reach of perceptually grounded symbols in accounting for the abstract amodal symbolism required in deriving the certain eternal truths of logic to which rationality *ought* to conform.

[2] In Gigerenzer (2002) a telling example of a heuristic is given comparing humans with robot ball catching. I believe that the example simply shows the normative strategy is falsely associated to this particular problem. One ought to catch a ball by keeping it in focus, not by calculating where it may land. The real normativity here is akin to the frame problem: what variables are (ir)relevant..

Adopting an implicit behaviorist position as per Gibson (2001) is methodologically more productive, in my view, as it avoids unnecessarily rationalistic "top-down" or "hypothesis-based" (e.g. Papathomas (1999)) free-for-all explanations. Nevertheless, whilst avoiding talk of representations is basic to what I will argue for here (see specifically Chapters 4 and 6), adopting this position does not suffice. In order to defeat the neo-rationalist account of a rational core in beings otherwise fully irrational, one will need to provide another plausible, but representationless, way to account for the human behavior of dealing with non-perceptual symbolic arguments.

If one adopts a rationalist position, with a cognitive core handling amodal symbolic argument, one will not be able to avoid the quintessential dualist issues of interfacing[3]. When, on the other hand, one adopts an exclusively perceptually grounded position one will, sooner or later, need to make a non-perceptual leap into amodal systems. Both views need some kind of mediating function in the form of common sense. As is clear from the word – at least in English – it connects the sublunary probabilistic via the term "common" and the eternal truth via "sense" (see below).

1.1.4. Context: the Freudian antinomy

Man consequently seems a highly schizophrenic – in the popular sense of the word – animal with both feet resting squarely in the context of everyday reality and a head high up in the thin air of eternal truth. This apparent basic human split personality has perhaps never been more clearly captured as in the Freudian dichotomy of the conscious and the subconscious. A dichotomy, which, as far as I can see, is not very far, removed from the recent proposals on cognitive architecture already mentioned above (see Stanovich & West (2000), Kahneman (2003) and Donovan & Epstein (1997)).

It certainly appears to be the case phenomenologically that there are two quite distinct "I's" at work within the head. On the one hand, there is an intuitive id that works its way from the ground up through the context of experience and that remains firmly attached to this context of contingent experiences. On the other hand, there is a firmly analytic super-ego that methodically crunches "Turing-style" whatever presents itself according to the pure laws of logic given "sub specie aeterna". That both of these ends ever meet within a single individual not only looks mysterious, it is mystifying. Indeed, the very same context that is a blessing for intuition and common sense is a curse for logic.

Working from the bottom up, assuming only an objective regular world around us, the mechanisms of probability (see below for a treatment based on Kyburg (2001)) allow a relatively straightforward way to reduce context as one goes along living in the world and gathering data. By simply living one can account for the forming of something like schemata (Cheng as per the review of K. Dieussaert (2002)) or related notions such as simulators (Barsalou (1999)) or scripts. The longer the period of learning and the higher the range of stimuli, the more adequate and far reaching one's world knowledge will be – and there is no species known to us with a longer learning period and higher stimulus range than the human race[4]. But that's about as far as it goes: perceptual experience can reduce context but it can't go as far as leading to context less eternal truth.

Working from the top down, we have every reason to be optimistic about "making sense" of the world given the social and scientific advances we have made. There is a host of evidence supporting us in the conviction that the world around us can be analyzed in terms of objective mathematical regularities. It also certainly seems that when we think consciously about the world, we think in such terms. It would indeed seem that we have some kind of "Turing" in our heads and, if so, that would most conveniently allow for an almost unlimited freedom for our ego's to account for the world in the mathematical terms that are such a hit in expressing and capturing knowledge about the world. Each human on that account is a mini-scientific community whose range of discovery is only limited by its limited life span and its differentially limited processing power. But that's about as far as it goes because, barring omniscience, there is no applying this knowledge in the context of everyday reality. "Given enough time", "ceteris paribus" and "in principle" may all be very sound and valid for axioms but they provide no clue as to a decision that needs to be made "on the fly".

---

[3] Including the frame problem as per section 1.1.1.

[4] I have always believed intuitively that the "poverty of the stimulus" approach was a travesty of an argument concocted by people that were deeply suspicious of the common sense of both children and their parents that child rearing is hard work. If there is an evolutionary miracle it is not consciousness or language but the fact that a species evolved spending the better part its life in learning and teaching.

Ultimately, the antimony allows us to focus on the issue of selective attention. Clearly, there has to be a means for selective attention to be driven by context as only in a certain context picking out (selecting) a subset of the facts relevant to a decision is possible. Without such context-driven selective attention it is quite impossible to avoid an eternal suspension of any decision. Nevertheless, at the same time there clearly needs to be a mechanism driving selective attention regardless of the specific context. If not, the application of the laws of logic, probability and science would be impossible unless one wants to posit some mystical way in which these laws are "directly" given as relevant context to humans.

Whether one starts bottom-up from perceptual grounding or top-down from analytical truths (or static hypotheses), the prima facie valid phenomena of the conscious and the subconscious seem respectively to overcrowd the same brain-space. Contextual probabilistic and pure logical constraints on rationality come together in the mediating function of common sense driving the process of selective attention. But, on pains of forever remaining in such a bottom-up vs. top-down antinomy, we need to post as an essential characteristic of this function that it mediates between something inside and something outside of the head.

1.2. A commonsense plan for reasonable housing:

We have found that traditional accounts in cognitive psychology lead to unsustainable psychological towers of reason. The babel of tongues that followed from this overly ambitious account of the powers of the brain directs discussion away from the basic fallacy contained in this towering reason. Below, I will first try to make the reason for this failure more apparent. After that, I provide a running metaphor that will, hopefully, serve to clarify the outline of a more modest alternative approach centering on the concept of common sense.

1.2.1. The fallacy of the psychological towers of reason:

In an, admittedly very informal, tentative and provisional way, we have seen that one can't expect pure rationality, whether bounded or nonmonotonic, from the single individual brain taken in isolation. Such a towering reason is in one sense much more than can be catered for by individuals whilst in another sense it just isn't enough.

Indeed, from the Fodorian antinomy it is clear that pure reason expressed computationally gets stuck in the frame problem. Alternative holist or connectionist theories on the other hand do not allow even explaining why such a frame problem would be expressed in the first place. As regards intuitions, it seems of no avail to fit human reasoning in a Procrustean bed of logic or vice versa. Similarly, assuming the basically Cartesian strong dichotomy between cognition and perception one arrives at the impossibility of such diverse phenomena being mediated within the same living entity. Finally, the issue of control on selective attention as captured by the Freudian analysis does not allow to be captured by accounts that start either purely from contextless Truth or exclusively context-sensitive Experience.

The often-implicit assumptions in traditional accounts in cognitive psychology can be labeled (to come back to the Kantian theme) as neo-rationalist (starting from the Truth) and neo-empiricist (starting from the Experience). Their failure lies in their shared ambition to account for all facts of reasoning in the workings of one individual brain. Nevertheless, their failure is not symmetrical. The neo-empiricist will start from something that is true and fail eventually because her ambition is to explain things that can't be explained purely on the basis of individual human experience. The neo-rationalist starts from what is false but his failure only becomes apparent in the Fodorian breakdown to account for the actual human behavior in computational terms.

The fact that languages are invariably structured recursively is the main rationale for the Chomskian agenda of looking for computational or purely rational brain functions (or Fodorian brain organs). But this fact need not be a fact about the working of our brain[5]. It can simply be a fact of languages, which is no more mysterious than the fact that a Newtonian description of physics – by and large – holds (nobody, I guess, would conclude from this that the brain, by and large, computes the trajectory of

[5] Not that this discredits any of the empirical work in this tradition as such, although it may discredit some of the theorizing specifically with respect to the link with "strong AI" and evaluative intelligence such as contained in Stanovich & West (2000).

objects based on such a Newtonian description). The initial success obtained in computational psychology probably explains why alternative proposals in changing the type of explanation sought have still retained a same functionalist (over)ambition. Without that ambition the challenges and quips of "zombiology", behaviorism and the like were all too probable. But, as said above, it is precisely this towering ambition that ultimately leads to the failure of both traditions.

This thesis aims at making it plausible that psychology or brain sciences taken as such cannot account for pure reason. It is futile to look for ways to implement deductive or other pure reasoning in neurons. We should instead focus on a more modest but maybe equally hard problem: how can the brain serve at the same time to get tuned to the physical world of things and to the rational world of words and of symbols. Such a conjecture is inspired by Popper (1979), Dennett (1994) and Jaynes (1976) – the present thesis however has the ambition of going one step further identifying the mechanisms (or at least their descriptive counterparts in pure reason) that can start to make good on the promise of the more modest but still reasonable housing afforded by common sense

1.2.2. A new plan for the house of reason:

I will attempt to make this conjecture plausible in the following steps through a metaphor. In lieu of a tower of reason contained within the brain criticized above, I will propose a house of reason where the social (critical, collective – see Popper (1979)) activity is an integral part of the explanation and serves as the common roof of pure reason. That roof obviously needs to be built from experience of individual human beings. Such experience needs to be able to serve as a strong foundation that is, for the essential part largely common, if we put to one side the specifics of cultural practices. Finally, the house itself needs walls and rooms to live in. These in turn are provided by the hybrid nature of human common sense that is grounded in the commonality of experience but allows to partake as well in the benefits afforded by the common roof of pure reason.

In Chapter 2, I talk about the house itself and the nature of common sense. I illustrate that the "if … then" conditional does not uncontroversially equate to the traditional truth-functional "material conditional" interpretation. This is done based on the work of D. Edgington (1995 & 2001). Reviewing prior psychological experiments, for instance on the Wason selection task, I hope to show that some of the previous interpretations were hasty. I also hope to show that this newer interpretation proposes a credible psychological bridge mediating (selective attention) between the roof of pure reason and Truth and the perception-based foundation of Experience.

In Chapter 3, I turn to the foundation. I attempt to show that, assuming an objective reality outside us that permits of a description in logico-mathematical terms, common concepts can be built as well as common choice and decision strategies adopted based on having a common experience. This is done with reference to the framework proposed in H. Kyburg (2001). The descriptive[6] validity of this framework as a proper normative framework of inductive reasoning is tested via a review of empirical results in choice and decision theory and categorization. The link to common sense and language will prove to be an integral part of the theory of uncertain inference adopted.

In Chapter 4, the roof is inspected. Although the inspection will be brief – since we are from here on literally beyond common sense – I hope to show how social interaction of commonsensical humans can explain the discovery of pure reason. I will do so by considering the work of H.P. Grice (1989), not as usually limited to his work on conversational implicature, but specifically as to the link between non-conventional (commonsensical) and conventional meaning. Again, a short review of empirical work in psychology should show how not making a split between social pure reason and common sense can be seen to lead to confused and muddled interpretations of experimental results.

After having given these three different parts of my conjectured house of reason, a comprehensive test for the over-all theory is proposed (but not executed) in Chapter 5, alongside a thought experiment. In a concluding chapter, Chapter 6, I briefly sketch how the emergence of an ever more sophisticated language as perceived by humans ("social perception") can give rise to complex phenomena as identity and consciousness. This will allow showing how this theory can connect with theories (such as Jaynes (1976) and Dennett (1994)) of consciousness as parasitic on the brains physiology, and morality.

[6] In line with the over-all thesis, I won't assume logico-mathematical theories are followed in the brain.

Chapter 2 – The "sense" in common sense:

"There's not much point in recognising that there's a predator in your path unless you also realise that if you don't change direction pretty quickly you will be eaten."

"On Conditionals", D. Edgington (1995).

We leave now the more poetic slumbering of the previous chapter to delve into some hard stuff. Hard empirical stuff delivered by experimental psychologists, for the present chapter specifically those with a focus on reasoning. But also some hard philosophical stuff, in this case regarding the interpretation of natural language conditionals.

The chapter has been written so as to allow separate reading of it. Section 2.1 basically summarizes the work of D. Edgington on conditionals and criticizes the unconditional acceptance of one interpretation of conditionals in the reasoning literature. Section 2.2 contains a discussion of how the proposed Bayesian interpretation outperforms other explanatory frameworks of human reasoning. In section 2.3 this Bayesian interpretation performs when benchmarked against empirical results (Dieussaert (2002) – henceforth KD – is taken as the guide in these sections). Finally, the link between this theory and common sense is explored in section 2.4. In doing so it will become clear how the material of the present chapter fits in the over-all "plan-of-the-house" metaphor, i.e. how it deals with the pivotal role between fallible, all too human uncertain knowledge and the critical, social enterprise of constructing theories of pure reason and scientific knowledge.

2.1. The edge of the "Edgington" conditional:

The classical normative interpretation of conditionals is truth-functionally straightforward. It is defined by a truth table where the conditional is true when antecedent and consequent are both true or where the antecedent is false, no matter the truth-value of the consequent (and where it's false otherwise). This interpretation – henceforth the "material conditional – is implicitly or explicitly used as the measure against which human reasoning performance needs to be assessed. Human performance fell dramatically short of that measure – a fact that was generative of a substantial part of the tradition in psychological reasoning research (Samuels et al. (1999)).

In this section, I will summarize a more recent theory for the interpretation of conditionals as given in D. Edgington (1995) and D. Edgington (2001) – henceforth DE1 and DE2, respectively. Dorothy Edgington developed her theory on conditionals in an article of "Mind" (DE1) that starts as follows:

"The ability to think conditional thoughts is a basic part of our mental equipment. A view of the world would be an idle, ineffectual affair without them. There's not much point in recognising that there's a predator in your path unless you also realise that if you don't change direction pretty quickly you will be eaten.

Happily, we handle ifs with ease. Naturally, we sometimes misjudge them, and sometimes don't know what it would take to be in a position to think or say that 'B if A', what would count for or against such judgments, how they affect what we should do and what else we should think. They cause us no undue practical difficulty.

The theory of this practice is another story. (..)"

This is in sharp contrast to psychological work in the area of conditionals. Maybe the sharpest contrast is seen in the work of Stanovich and West (2000) where it is rather the practice that is another story. Nevertheless, Edgington's sober stance with respect to theories – plural – on conditionals seems to be rather well supported by the facts. The relevant sections of KD clearly show a picture of a variety of interpretations of conditionals where it is, to say the least, hard to find fault with human intuitions. Sticking to a theory at odds with these facts maybe, to paraphrase D. Edgington, "an idle, intellectual affair" misapplying the standard of certain apriori reasoning to a contingent everyday reality of coping with uncertainty. It may be worthwhile to highlight the "ecological" context of her example. If, as she

argues, it is not the material conditional that describes the practical interpretation of the conditional, then one should be open to that practical interpretation not being exclusively human. Thereby we could have an element for an evolutionarily plausible account for emergence of pure reason (see section 2.4).

However that may all come out in the end, let me again quote D. Edgington to scope the present exercise with some more care (see p. 390, DE2, after describing counterintuitive results of the classical material conditional and the possible truth-functionalist defense in the face of that):

"(..) Natural language is a fluid affair, and theories cannot be expected to achieve better than approximate fir. Perhaps, in the interests of precision and clarity, in serous reasoning the untidy and unclear 'if' should be replaced with its neat, close relative, 'Material Conditional[7]*'.*

This was no doubt Frege's attitude. Frege's primary concern was to construct a system of logic, formulated in an idealized language, which was adequate for mathematical reasoning. If 'Material Conditional' does not translate perfectly our natural language 'If A, B', but plays its intended role, so much the worse for natural language.

Perhaps, for the purpose of doing mathematics, Frege's judgment was correct. (..)

The oddities however become less tolerable when considering conditional judgments about empirical matters. (..)"

There will be those, no doubt, objecting against such an explicitly acknowledged dichotomy between mathematics and natural language. It would be unsurprising if those objecting would, roughly, coincide with those appealing to pragmatics as reason for what would, on such a reading, *only* be a pragmatic gap. However, she (p. 391 ff, DE1) explicitly addresses this challenge straight off the bat. Assertion or conversation is not necessary preconditions for the oddities to occur. One need not assert or express one's conditional belief for that belief to be at odds with classical readings of conditionals. To give an example out of DE1, p. 392:

"Thinking you won't eat the mushrooms, I may without irrationality reject

If you eat them you will die."

Where the underlined conditional is true, under the classical material conditional interpretation, for the mere reason that its antecedent is false.

All of this highlights two more implicit assumptions of D. Edgington's final thesis that are specifically relevant in the context of this thesis:

- conditionals, in practice, are not used to express certainties but rather conditionals allow us to frame regularities without needing to be explicit on the "certainty" involved, and,

- conditionals are not primarily a device for communication nor even for internal reasoning, they rather are a means of "effectual" bookkeeping of uncertain knowledge.

Before proceeding, it may also be useful here to highlight that conditionals are not fundamentally about causal links nor about theories or explanations of how the world is or needs to be. Conditionals serve their use adequately by tracking consistent correlations. Whether or not these regularities are based in causality is already beyond natural language and psychology and well into "pure reason" metaphysics.

With these preliminaries out of the way, I can present the basic interpretation of the conditional as it is given by D. Edgington (DE2 p.263, I use this formulation with "b" standing for "your degree of belief in" to avoid confusion with the work on probabilities by H. Kyburg (2001) addressed in chapter 3):

b(B if A)=b(A&B)/b(A)

In other words, this conditional expresses the belief that B is considerably more likely to occur than to not occur when A occurs. This can be graphically demonstrated by considering beliefs to be structured

[7] The corresponding logical symbol is used in DE1.

as probabilities and by drawing the partition of states relevant in the above formula (see DE1, p. 397, Figure 17.1).

Assessing "B if A" in such a drawing basically boils down to comparing the relative sizes of "A&B" vs. "A & not(B)". In the above case the odds are 5 to 2 and the conditional probability of "A" is 5/7. As one can immediately see the formula can only be used when the "epistemic possibility" of A (b(A)), is non-zero. I will not go into this and other consequences of the theory into counterfactuals, the reader is referred to DE1 for an introduction and DE2 for a complete treatment. The above related to indicative conditionals where A is an epistemic possibility is all I need for the purposes of this thesis.

On this practical interpretation of conditionals, one expects that people will assess their experimental task involving conditionals (and conditional beliefs) as if they were assessing the Bayesian expression corresponding to it according to the above formula and graphical representation. This does not mean that experimental participants compute – in any meaningful sense of that term – Bayesian statements. It only means that they behave (reason), roughly and intuitively, in ways best described in probabilistic terms. The "roughly" is key in the sentence as nothing in the foregoing requires precision, nor does anything in the foregoing expect precision. On the contrary, such an account is profoundly adapted to a "logic of uncertainty", which is called for in both Kyburg (2001) and DE2, and fallibility of judgments whose precision can only be improved by experience (see Chapter 3).

It is with some hesitation and unease that I cut this story short without considering the broader scope as is addressed in both DE1 and DE2. However, I hope others would be tempted by this to make a more elaborate review along the below lines. I am sure that that specific journey will be a rather productive one allowing a high level of systematicity (see, for instance, the Table on p. 401 of DE2).

2.2. Conditions on psychological theories of conditionals:

Traditionally, psychological theories tried to explain the divergences between the empirical results and the normativity of the classical material interpretation of the conditional as a matter of "performance" – in a way closely analogous to the use of that term in the linguistic field inspired by N. Chomsky. The untenability of this interpretation is argued for in section 2.2.1 below. Refuting the traditional accounts is, however, not sufficient for my purpose. Therefore, in section 2.2.2, I will need to go one step further in criticizing some early probabilistic theories which, however much closer they are to my present conjecture, are still in my view basically mistaken in both their ambition and their means.

2.2.1. Mental muddles[8] and unruly mentality:

It was E. J. Lowe ((1993), p. 211) who questioned the basic approach of mental models (see review in KD, p. 41 and Johnson-Laird (1993)/(2001)) with an argument aimed at any explicit human reasoning ability following the formal methods established in the province of logic. He said:

"Boole and Frege represent fundamentally opposed traditions concerning the relationship between the science of logic and the psychology of reasoning: the former holding that the empirical study of mental inference reveals the content of logical laws, the latter that logical laws are a priori principles which provide norms for human reasoning rather than descriptions of it. The view that I shall be defending runs counter to both of these traditions because it denies that our basic logical competence consists in an ability to employ general logistic methods of the kind studied by formal logicians like Boole and Frege and their modern descendants. Such methods provide only a variety of techniques for extending our powers of inference beyond their normal range, and as such are neither revealed by how we do reason nor constitutive of how we should reason in basic cases."

Although I would go further than E. J. Lowe, not being afraid to disagree with Locke in that God left it to Aristotle to make sophisticated creatures like us rational, the basic gist of this critique is similar to

[8] The term was (Kristien Dieussaert [personal communication]) coined by Rips (1986).

mine. I believe however that with the aid of what was said above, the criticism can be formulated more directly. In a way that moreover allows to make definite predictions of new phenomena as well as the unified accounting for various already established phenomena (as attempted in section 2.3).

In order to start with this, let us take as example the different treatments for modus ponens (MP) and modus tollens (MT) inferences according to the mental model theory (a summary can be found in KD, p. 41). The fact that MT is more difficult and more error-prone (if judged by the classical reading of the conditional as a material conditional) is deemed to be explained by the mental model theory by the fact that MT requires more model building than MT.

However, on the above reading of the conditional, the conditional "If A, B" reduces to the assessment of the two leftmost parts of the below partition (figure repeated for convenience):

As is immediately apparent from this graphical representation, the relative size of the two leftmost parts of the drawing is what makes or breaks the "truth" of such a conditional. These parts are exhaustive as to what happens "in case A". In fact, I do not think that there is a real difference in natural language practice between the conditional and its corresponding MP inference. The relevance of the conditional is the relative strength by which MP holds (on a subsequent assertion of "A" interpreted as per D. Edgington [personal communication] as an assertion of p(A)=0.9, p(B) is proportional to p(A/B) – but see section 2.2.2 and Chapter 3 on why an interpretation like that is, at least psychologically, problematic). I will refer later to this as "the surface structure" of "If A, B" is such that it immediately allows to pick out what's relevant "If A", namely B.

In the case of MT however – in asserting "not(B)", the surface form of "If A, B" however makes us none the wiser. Indeed, there is nothing in the above graphical representation that gives us a clue on "in case not(B)" (none, all or "anything in between" of the rightmost "not(A)" area may be "not(B)"). Let me illustrate the point by providing two partitions consistent with the partition above explicitly vindicated on Edgington's theory for "If A, B":

Both of these partitions exhaust the possibilities of "B" and "not(B)" in entirely different ways but they are both commensurate with the direct interpretation of "If A, B". In other words, it should be obvious that the original conditional "If A, B" does not provide clues as to which specific shape of partition (or for that matter which shape of the infinitely many options comprised between the two given above) is the correct one. Despite the logical consequence of both the mental model and the mental rules (Rips (1986)) accounts, the conditional "If A, B" does not provide us access to what happens "If not(B)" (see also DE2, p..401 argument family (6)[9]). The best one can do on hearing not(B) is to conclude that it is

[9] It is this kind of table that would provide a very concrete and very productive means of interpreting a host of reasoning experiments. Certainly if combined with the methodology of treatment suggested in DE2 and applied above via graphical representations of partitions.

not very likely to happen in case of A which is not the same as inferring from not(B), that not (A). In another semantics, there are at most a few not(B)-worlds that are also A-worlds but how many not(B) worlds are not(A) worlds rather than A-worlds is basically undefined by the original conditional.

As has been pointed out to me by D. Edgington herself [personal communication] this does not mean that MT is an unsafe inference to make. On the contrary, taking the assertion "not(B)" semantically as p(not(B))=0.9, MT will ceteris paribus (i.e. with p(B/A) fixed in both the MP and the MT case) yield a higher probability for "not(A)" than MP will for "B" (see also section 2.2.2 immediately below). This being as it may, psychological experimentation is hardly very clear on what is asserted and to what extent it is asserted. To all extents and purposes the traditional accounts treat MT equivalently to a type of inversion to "If not(B)" which is basically invalid on this interpretation of conditionals.

Within the view expressed here, human beings do not even make the basic MP or MT inferences and work directly from the surface form of conditionals to whatever the "relevant" consequent is. The short explanation of the difference in difficulty between MT and MP according to this thesis is that MP is the inference "aligned" with the basic conditional expression whilst MT isn't. Anybody "dealing correctly" with MT is doing so for reasons perfectly foreign to the concept of MT.

There is obviously more work to be done here. For instance in reviewing specific experiments along the above lines to show that "predictions" as conceived in both the mental models and rules traditions are nothing else than arbitrary post-hoc interpretations of basic phenomena predicted by the adoption of the "Edgington" conditional. As pointed out in E. J. Lowe (1993), the explanatory burden of theories as these is mostly carried by ad hoc assumptions on "performance", e.g. which models are not constructed or which method of proof is used. It would also be of value to address directly the mental rules accounts (KD, p.49 ff.) but I believe I can rest my case here on the basis of E.J. Lowe's convincing arguments that there is and cannot be any real principled difference between an account based on formal syntax and formal semantics (except, of course, for a potential arbitrary move of making mistakes or not computing certain models that at any rate should be able to explain any effect or "fallacy").

2.2.2. Probable shortfalls of probabilistic explanations:

The type of problems encountered by syntactic and semantic theories of psychological reasoning have already prompted many to look for a more probabilistic explanation of human reasoning. In fact, E. J. Lowe – in the paper referred to already a number of times – explicitly suggests such a solution, as is done in that same book in Oaksford & Chater (1993), a summary is given in KD, p. 51. The difference with my conjecture is that I radically maintain that humans do not perform calculations or computations for the inferences they make. Whether or not the calculations are probabilistic or not does not matter.

My view is quite probably at odds with D. Edgington's view. She refers often to E. W. Adams, e.g. to E. W. Adams (1975), and the concepts of supervalidity and probability preservation that could well be seen as candidate principles underlying probabilistic reasoning. Nevertheless, as she herself remarks, (DE1, page 267):

"Conditional degrees of belief is an interesting concept to the extent that the ratios are stable fixtures of a belief system, which can be settled independently of b(A) and b(A&B)."

In other words, for a conditional to hold one need not quantify the probabilities involved. Rather, when one assumes calculation of probabilities (whether with "connectionist" calculators or not), one can see from the above example in section 2.2.1 that MT is actually "safer" than MP. If people would calculate with probabilities in order to make inferences, this is directly at odds with the basic observation that in practice MT is "more difficult" for humans than MP. Leaving out any inference, as I propose, is then a simple alternative that accords better to the facts.

Consequently, I do not believe that a theory of conditionals provides a theory of assessing conditionals, nor a theory of constructing conditionals. In a Gibsonian way, I see no reason why people (and animals for that matter) would not formulate their knowledge in conditionals without any direct grasp of the underlying probabilities, or anything else than picking up directly regularities in the world. These conditional beliefs may well require addition of simple (categorically asserted) probabilities in order to

allow explicit inferences (as much is the gist of D. Edgington [personal communication], see above), but this shouldn't be confused for what's happening cognitively with conditionals. The former is an exercise of pure reason whereas the latter is more an exercise of nature that can be merely described in the terms taken above. Nature is very allowing in its subjects evolving an adapted behavior over time so the degree to which "If A, B" holds can be as fine as possible without any explicit computation assuming only that "If A, B" as adopted here plays its role in survival.

This does not mean that people would not be sensitive to arguments that are suspect on probabilistic readings of argument strength. It only means that their sensitization is not the result of computations as to invalidity but rather the result of experience with the fact that such suspect arguments yield suspect results. By the same token intentional strategies such as Optimal Data Selection (Oaksford & Chater as reviewed in KD) should not be explained as inherent psychological competences. These strategies, on my view, are:

- conducive to survival, i.e. its corresponding behavior can be copied (along the lines of "If behavior, then success") as a higher order invariant regardless of any underlying knowledge as to the soundness or (super)validity of that behavior (a similar argument is made in KD via the related concept of "conditioning"), or,

- the result of explicit instruction, i.e. conditioning certain conscious behavior using the limited capacity of working memory as a "note pad" on which to reconstruct the strategies of "pure reason" in the absence of writing tools and social criticism.

I will come back on the latter point in chapters 4 and 6. Suffice to say here in this regard that in that limited sense all of the above - mental models, mental rules and mental probabilities – may be adequate descriptions of reflective conscious human behavior. The latter consisting quite literally in playing out critical, social interaction within working memory akin to "remembering" the carry digit in summation.

As to the former point, I believe it is not productive to adopt the view that any of the above are true explanations of reasoning just because they happen to describe arguments that are, analytically, true. As none is in any non-exceptional way the explanation of how we come to infer things. As already noted before postulating embedded competences is an easy way out, if one – as is unavoidable, see chapter 1 – then postulates post hoc wild cards of "performance" to explain the divergence with actual empirical results. The Chomskian concept of resolving a surface form into a deep structure allowing analytical treatment via models, rules or probabilities is paradigmatic for these approaches.

In this sense what I'm attempting here is closer to the project of D. Wilson & D. Sperber understood as relying entirely on what is given as surface structure to explain "reasoning" in humans. On such a view, arguments that are supervalid in a probabilistic sense are so quite independent of our brain circuitry. In a sense we can ***show*** their validity by constructing externally to our brain underlying deep structures. We pick up the fact that "if" a surface form (maybe we can call them logical forms) corresponds to a certain deep structure analysis "then" we can expect a corresponding inference to be safe. Externally to the human brain deep structures can be found for anything deep structures were hypothesized for within the brain. The above Edgingtonian analysis of the conditional allows to capture such regularities and as such allows a direct link to pragmatic reasoning schema's to which I turn next.

2.2.3. Schema's are conditional upon this conditional:

Pragmatic reasoning schemas are indeed the proposal on record that is closest to mine. I will consider it here with reference to the treatment in KD (*), p. 116, Manuscript 3. The following quotes are part of the conclusions from experimental items roughly related to the above, interspaced I provide relevant links to the Edgingtonian interpretation of the conditional discussed above:

"The most specific theory in terms of the content question is the theory of the 'pragmatic reasoning schema's'. Cheng and Holyoak (1985) explain the context and content influences on inferences by pragmatic reasoning schema's. They state that in realistic situations, people reason with pragmatic reasoning schema's. These schema's find their origin in the abstraction and generalization of specific, personal experiences."

As will be clear from the above, I am very sympathetic to this view. I propose to call this process of abstraction and generalization "social perception", referring to the fact that in my view social practices of reasoning get picked up this way (in a way strongly parallel to visual perception) and so form a basis for forming schema's.

There is, nevertheless, a substantial difference between the conjecture here proposed and schema's of the type proposed by Cheng & Holyoak (1985) or the, in my opinion, similar concepts of simulators (Barsalou (1999)), scripts and the like. Schema's are considered as enabling conditions for conditional inferences. My proposal on the other hand considers conditional inference – which, remember, I do not see as distinct from the conditional itself – as the basic mechanism enabling pragmatic schema's. I am well aware that this statement is open to challenges of being only terminological but I hope to show this difference really is crucial (as is maybe immediately apparent by the fact that with my proposal the link between objectively (super)valid arguments is guaranteed along the lines at the end of section 2.2.2.).

To make a start with such a demonstration, let's consider the following conclusion of KD:

*"A minor point (*I understand what is meant is: a weak point – Jo Bervoets*) of this theory, however, is its explanation of the 'competence question'. Evans (1991) stressed that a good theory should account for three questions: competence, content and bias. Surely a good theory is one that does not loose track of the other two questions in trying to account for one question."*

It should be clear that I resist using the terminology of "competence" as it leads one easily astray into a computer analogy where competence is interpreted as a discrete faculty of reasoning following logical rules (whether syntactic or semantic). This being said, I do think the remark is pertinent as the theory of pragmatic schema's (and so do its brethren concepts) suffers from lack of detail in accounting for how they're formed. Reference to associationary mechanisms – connectionist mechanisms –does not, in my view, suffice to account for forming and schema's. Roughly, the theory is open to challenge because it does not seem to be true that forming a schema is based solely on it having been perceived often – most people would accept that there is such a thing as one-shot learning. Similarly, activation of a schema does not seem to be based exclusively on the broad context or environment that constitutes the cue – the analogical use of a schema outside of personal experience is enough of a challenge for the present account.

What is lacking then, in my opinion, is indeed addressing the competence question – or the question of common sense. I believe that this question can be addressed working bottom-up from a "competence" of forming conditionals embedded in a comprehensive logic of uncertainty as provided in chapter 3. I believe that such proto-schema's will allow the subsequent proliferation and sophistication of schemas in response to increased sophistication of reasoning ability in society. But a full treatment of this aspect will have to postponed until after chapters 3 and, partly as well, 4.

2.3. Empirical conditions to be met by this conditional:

I will now try to demonstrate how a variety of empirically observed reasoning phenomena involving conditional statements can be directly explained based on the above theory of practical interpretation of conditionals. I will be following the outline of KD for ease of reference and clarity of exposition, citing the page numbers as well as the original source publications involved. This can be nothing more than a sample overview showing the methodology used in interpreting these experimental results in a unified way based on DE1 and DE2.

2.3.1. Reasoning in semi-conflict:

As per KD, p. 56, an experiment set up by Byrne led to the following result:

"On presenting two conditionals with different antecedents but identical consequents, MP is suppressed after the categorical assertion of the first antecedent."

Such suppression is perfectly in line with DE2, p. 401 item (4), "strengthening of the antecedent". The addition of an antecedent makes it unsafe to infer the consequent.

The question remains of course whether or not and if so, why, the two conditionals are interpreted as a strengthened antecedent. This is, at least on my account, a risky question as I need to answer it without reference to any psychological construction of such strengthened antecedents, or other computations or calculations, whether or not of a probabilistic nature.

For instance the proposed criterion of semantic relatedness is a slippery one as it imports context and, by importing context, a potentially intractable solution running into all kinds of frame and qualification problems.

In the present case, I believe the answer is straightforward. If strengthening of the antecedent is known by experience to be suspect (maybe a better term in this context than invalid) then one only needs to look at the surface of the two conditionals with identical consequents to trigger that belief (which can clearly be registered as a conditional itself "If an argument of the form SA, then MP is suspect").

Alternatively one can draw the partitions corresponding to the two conditionals and observe that any combination of the two conditionals can have unforeseen effects with respect to the first antecedent. Again, however, I need to stress that I do not propose such a process to have psychological validity. I merely believe that such a graphical illustration of the facts of the case explain why it would be natural to assume that people's experience is commensurate with SA being suspect.

Note also that the desideratum of semantic relatedness rather than being seen as a cause, can on this view be seen as a consequence. Indeed, if the second conditional is unbelievable, it should not trigger a "suspect" SA. If, on the other hand, it is believable but its consequent is somewhat different it shouldn't have the SA effect. Inverting causal roles is on my account crucial to solving the mystery of common sense.

Obviously, the effect need not be limited to when conditionals are explicitly stated. A first conditional, by its specific surface form may well pick out a number of related conditionals by association. Whether or not these associations are made – in line with the thesis of this thesis – does not depend on any other process than previous experience. A culture that never spends any relevant time inside will never find a second conditional relating to being inside.

This is in fact a first explanation of common sense and more specifically the common part of the phrase as people's dealing with such conditionals will be based on previous experience and learned wisdom, in the form of conditionals and corresponding associations between certain surface forms.

2.3.2. A proposal for peace in the war on intuitions:

The putative wisdom expressed in the last paragraph is discomforting to many (but see KD for similar intuitions and related research) as it denies a psychological role to formal methods of reasoning. Even worse than being infected by mere context, it would lead to a view of human reasoning that is almost fully determined by experience and association. Small wonder then that many researchers have naturally tended to use material for which there is little prior knowledge disturbing the picture of the underlying reasoning process[10], see also – for a summary of that type of research – KD, p 59 onwards.

The best and most tenacious examples of this research can be found in Ford & Billington (2000), Ford (2004) and Elio & Pelletier (1997). Although these lines of research are not exclusively dealing with conditionals, I believe it worthwhile to treat them here within the broader context of "probability preservation" reasoning suggested in both DE1 and DE2.

I will not be the only one for which this work strikes me as the search for a needle in a haystack. However, I have not seen too many who would affirm – like I would – that the search is fatally flawed because in this case there isn't even a needle (rule) to be found in the haystack (empirical results).

[10] It is nevertheless not clear that in experimental situations, participants will use their prior knowledge. It's not unknown to most participants that in psychological experiments one is usually not interested in one's specific background knowledge. I will contend later that such effects explain anomalous results in some studies, the experimental subjects basically only using their stock of reasoning shortcuts, rather then using their stock of contextual knowledge. Most subjects can be expected to differentiate between syntax and semantics.

The underlying intuition of this "psychologistic" research can maybe be best paraphrased as:

"(..) instead of human reason to conform to logic, logic to conform to human reason."

Roughly, these psychologists are somewhat disheartened by the amount as well as quality of guidance provided from the field of logic. As to the former, whenever one goes beyond classical first order logic one finds an enormous amount of logical formalisms, each vindicated by specific intuitions and/or application domains (for an overview, see Gobles (2001))[11]. As to the latter, specifically the nonmonotonic logical formalisms that are developed to deal specifically with the types of reasoning problems in which humans, prima facie, excel seem to be of a staggering psychological implausibility (see Ford & Billington (2000), specifically quotes on p. 446).

For this reason, one tries to eliminate the effect of experience using methodologies akin to Osherson (1992) in deriving experimental items that both eliminate the potential for solving reasoning problems by some direct appeal to experience (see Ford & Billington (2000)) or limit the effect of context to a well controlled experimental set-up (sci-fi or other cover stories, see Elio & Pelletier (1997)). If certain formal mechanism for solving reasoning problems were present, they would be made salient by such a type of experimental items. This is, I believe, the implicit assumption taken in this line of research and it is one that is correct. So, let us check the results obtained in a sample of these studies.

First, let me quote one of the findings of Elio & Pelletier (1997), p. 50:

"(..) the preference of our subjects to deny the conditional as a way of resolving contradiction can be interpreted as a preference to retain the truth value of 'data' (the non-conditional sentences) and deny the particular interdependence that is asserted to hold between them This seems straightforwardly rational from an empiricist viewpoint: the 'regularities' are nothing more than a way of summarizing the data."

In other words, the only consistently recurring strategy in dealing with these problems is one of giving preference to what is considered to be data. Reasoning phenomena can't be isolated or dissociated from the basic mechanisms of induction. Hence, human reasoning cannot be isolated or dissociated from experience. This is, at least for somebody holding that there is in some way a psychological faculty of logical reasoning, an outcome much worse than mere dependence on context. Indeed, the only pattern in depopulated experimental items seems to be something "extra-logical", i.e. counting on experience or being psychologically committed to the law of induction (see chapter 3 for principled "logical" ways to account for this fact).

On our reading of conditionals, nothing more nor less than this outcome could be expected. Conditional statements weigh the occurrence of B on the evidence of the occurrence of A. The reasoning process is not such that formal mechanisms based on manipulation of certain forms of argument come into play at all. Rather, the phenomena can be explained with reference only to the brute force of experience, for a direct experimental corroboration of this thesis see also KD, p. 104 (from an experiment where initial conditionals were given strength through experience and/or context):

"The data clearly show an effect of initial belief state (certitude level) on the answer pattern. In particular when people are initially more convinced of the truth of a conditional rule, they tend to retain their belief in the rule, even when contradictory information is presented."

In other words, conditionals record regularities of the world. There is more to a conditional than can be forgotten on the basis of singular counter-evidence nor. There is a conditional belief populated by quite some experience. Nor can a conditional established on the basis of a singular assertion blow away the belief basis formed through experience. Building and using conditionals takes experience, but once built and established their use can account for a great variety of phenomena of commonsense without recourse to manipulation of formal logical concepts. For more results in the same line see KD, Manuscript IV, where multiple items of relevance in the type of experience and background are investigated in more detail.

---

[11] The story goes that in an international conference for logicians a quite attractive logician sat next to an equally attractive other logicians. They happened to be sexually compatible. When the speaker on duty asked the rhetorical question "How should we prove this?", the former attractive logician softly whispered in the latter attractive logician's ear: "Your logic, or mine?".

Turning now to the research of Ford, it would seem that a number of recurring factors are identified in her research. A summary of these factors can be for instance found, tabulated as positive and negative factors, in Ford & Billington (2000), p. 459 and Ford (2004). From the former publication we take the factor labeled (N3) – within the scope of this thesis we can only take a sample but the method used here should apply to the other factors proposed by Ford:

"Some participants considered path length regardless of the ordering of rule types. Most participants preferred the shorter path to the longer path."

Whereas such a factor may be seen in the context of a specific experimental item as "negative", this is not as straightforward a labeling as it may seem at face value. Indeed, it is generally true that the longer the path of any argument the higher the likelihood of such an argument not being valid (in the sense of not preserving probabilities, see DE2 definition of supervalid). The best extreme example of this is the Lottery's Paradox (see DE1) where extremely long path length invalidates even the "safest" argument.

Without a sophisticated (put more directly: without any) faculty of reasoning it would be expected to dismiss arguments based on the surface form of their path length. The notion of social perception that was briefly introduced above would yield something like: "humans perceive that arguments with long path length are weaker based on their social experience telling them these arguments are suspect". In a sense of the word "see", one could say that humans "see" that some arguments fail to preserve enough probability before anyone "knew" that probability preservation was important. In this sense as well one could say that common sense, as tradition would have it, is a precursor of the formally proven "right" solution. But the issue with this metaphorical language is that it easily confounds people as to why this is so. Not because our brains have a sixth sense for these things, but because social practice has evolved to a point where arguments of this and other types, together with their objective success rates, are a real integral part of human experience. Instead of being positive or negative factors, all of these factors are merely lines of reasoning that are current in social practice and that, again through social practice, are associated to weakness or strength of the argument that is phrased in this way[12].

If this conjecture proves anywhere near correct, there are immediately several cautionary remarks to be made with respect to reasoning research and the much-debated status of intuitions:

- First and foremost, the good old Fregean idea that intuitions borne out by psychological reasoning research do not provide a good basis for developing, testing or validating systems of logic. Indeed, such research will uncover generally accepted – commonsensical – approaches to reasoning problems rather than any well founded logical way of dealing with them.

- At the same time, philosophical intuitions should not be suspect merely because humans don't seem to reason according to them in specific cases. The measure of philosophical intuitions is not their explicit social acceptance but their resistance to philosophical criticism and their use in practical applications and argumentation. The role of psychological research is however not reduced to zero. If certain mathematical formalisms are proposed for the logic of uncertainty, these formalisms can't run counter to psychological practice - the latter needs to fall within the descriptive idealization of the former. This is exactly what I intend to do here: rely on such a descriptive formalism to explain the behavior labeled "commonsense". The descriptive fit of formalism and behavior is not due to the latter instantiating a computation of the former but simply due to the fact that both are vindicated by a same external reality critically perceived.

- Although this concept of social perception can easily be mistaken for a concept of heuristics – or bounded rationality – such confusion is not warranted. If a certain type of reasoning pattern (logical form) is deemed weak or strong, then this is based on its status as socially perceived and, ultimately, its success rate in a process of trial and error. Argument strength for single individuals' common sense does not depend on theoretical insights. Much of common sense is dead wrong and, unfortunately, nonsense has become the norm of common sense in many dark periods of history. One of the deeper truths of common sense therefore is not to trust the

[12] A strong way of operationalizing this immediately suggests itself: varying arguments by presenting them in certain syntactic or surface structure ways should have a predictable effect on the level of strength that is associated to them, regardless of any content and context effects.

arrogance of common sense to give an adequate – let alone correct – answer, without putting it up for social criticism.

- The types of patterns uncovered by Ford, are clearly syntactic and speak against any kind of "pure" connectionist, associationist or probabilistic account of these facts if those would refer only to perceptually grounded symbols (in the sense of Barsalou (1999)) as the cornerstone of phenomena of "higher cognition". Syntax is the most direct way in which we perceive socially and should therefore have a causal status for human reasoning irrespective of any need for perceptual grounding. Any perceptual grounding of syntactic features lies simply in the fact that syntactic surface structures are perceived as syntactic structures in much the same way as visual perception derives higher order invariants relevant for e.g. object recognition.

It is important to note that Ford actually comes pretty close to a line of argument as developed here. In Ford (2004), p. 97, Table 4, she gives the results obtained by Gillio (2002) on probability preservation of some arguments traditionally considered in the nonmonotonic reasoning literature (see for instance for a recent overview, H. Rott (2001)). However, contrary to my conjecture, she (and E. W. Adams (2002) himself, it would seem) uses the information to set up an even more sophisticated set of rules ignoring the immediate feedback of her own protocols, where a majority of people constantly referred to "some likelihood", "probability" and generally taking the arguments at face value rather than setting up any an elementary reasoning as e.g. provided by e.g. Table 6 and 7. In doing so, she might get descriptively closer to the actual empirical results but at the expense of psychological plausibility. She might well find that such rules, if ever established, vary with cultural progress; a result that would be quite at odds with anyone's intuitions, I suppose.

2.3.3 Pragmatism afforded via pragmatic types:

I am certainly not the first arguing for a relationship between (conditional) probabilities and practical interpretations of conditional statements. As much is clear from KD, p. 70 ff., Manuscript I where Newstead et al. (1997) is quoted to argue for just that. Quoting KD, p. 71 (at the bottom):

"They (Newstead et al) argue that the differences between the pragmatic types are due to the fact that real life conditional statements normally express uncertain or probabilistic relationships rather than certain ones."

KD, Manuscript I provides a detailed, quantitative account of how certain variables (speaker control, alternatives/disablers and pragmatic type) covarying with (conditional) probabilities influence response of participants to conditionals in which these variables are varied. Without doubt, these quantitative measures should be directly and readily interpretable within the framework of the "practical conditional" introduced above. Specifically, as can be seen from sections 2.1 and 2.2.1, on our present hypothesis, minimally, b(B/A) and, maximally, b(A), b(A&B) (see the above formula and partition drawing) should be of influence on assessing such conditionals.

The present conjecture is therefore more restrictive as to the important variables as the environment of quantities considered in KD, specifically when adding the considerations of Note 1 as to the need to rely on the likelihood of (degree of belief in) B[13]. As far as I can judge, the results contained in KD are only superficially at odds with my more restrictive view as implied by the Edgingtonian conditional combined with the lack of computational processes on probabilities (see section 2.2.3 for a parallel with the treatment of pragmatic reasoning schema's).

First, L(A/B) – the likelihood of A given B, corresponding to b(A/B) – should not factor into responses to conditionals. I believe the results in KD cannot be construed as implying the contrary. As far as I can see, it is precisely with L(A/B) that this empirical work comes apart from previous research heavily biased by a classical interpretation of the conditional as a material conditional. This obviously does not mean that within this or that range of conditionals L(A/B) would not covary with the interpretation of the conditional. Indeed such covariance can be expected for conditionals that are conventionalized as requiring a material conditional interpretation (e.g. in logico-mathematical pursuits).

[13] I apologize for using the notions of degrees of belief, probability and likelihood interchangeably and therefore rather loosely. I believe however that the risk of confusion is limited in this context.

Second, one need not rely on explicit probabilities to account for the influence of pragmatic type nor of speaker control as recorded in this study. As I argued before, one can rely instead on social perception (see section 2.2.3) of certain pragmatic types as being associated with certain conditional probabilities (and therefore also of certain conditional degrees of belief). Specifically for the factor "speaker control" it is immediately apparent that the conditional probability of B given A is so much the higher if the one assessing A has higher control over B. Interestingly, such a fact of experience does *neither* require us to state some or other corresponding logical fact *nor* does it require to search an indeterminate amount of knowledge of the world.

Consequently also, there is no need to postulate "pragmatic" ways of finding "relevant" context. The surface structure of the conditional – in the present case one with a certain level of speaker control – suffices to pick out the relevant facts. In the case speaker control the surface structure will be of a form "If S wants B, B". We all know that with such a structure L(B/"S wants B") is high and corresponding to that, through social perception, we will come to be confident that b(B/"S wants B") is warranted. As mentioned already several times the (logical) form of "If S wants B, B" can be picked up without need of postulating intermediate representations and computations.

As a final note and by way of pointing out the ease with which certain misunderstandings can come about, I want to refer to Note 1 of KD, page 95. There it is suggested that L(B) is a factor that needs to be taken into account. This, if true, would refute the descriptive appropriateness of the Edgingtonian conditional and consequently my project. However, as suggests in KD: *"Three of the four threats were considered exaggerated (..)"* i.e. the issue is not with L(B) but with the entire conditional being unbelievable or L(B/A) being low. This is clearly borne out by the fact that a consequent such as *"get out of the car"* is not in se unlikely; it is merely unlikely in the context of A or, in other words, the conditional is unlikely to hold. Or, finally, b(B/A) is low.

Although out of the scope of the present exercise, a factor not mentioned but of considerable potential importance for interpreting conditional statement is b(A). Its potential importance is clear once one is to vary it in the partition depicted above. Some examples of its relevance can be gathered from DE1.

2.4. Conditionals, or the "sense" in common sense:

I have started planning the house from the middle-out. So much is common practice in planning for a house since, after all, one does not live in the foundation nor on the roof. But, after reviewing briefly whether the house itself provides enough space to accommodate us – with all our furniture in the form of empirical and a priori knowledge –, I should quickly turn to its foundation and its roof. A house that has enough room but which would fall down following the slightest critical tremor or would require ad hoc measures of repair after every storm for lack of a decent roof isn't much of a house, even if it looks almost as beautiful in your dreams as did the functionalist tower that failed to support itself.

Let's then first see what, on the above construal of conditionals, can be learned about common sense.

First, I hope to have demonstrated that the Bayesian construal of the conditional as is proposed in DE1 and DE2 is a plausible unifying theory for a crucial domain in reasoning research, where up to now an abundance of theories are pitted against each other (each with an abundance of post hoc extensions). I also hope to have made it plausible as a purely descriptive theory, not purporting to explain the exact mechanisms underlying such a behavior. Indeed, my use of D. Edgington's theory is an attempt to account for reasoning phenomena in a bottom-up fashion. The Bayesian construal, when taken together with a probability preservation approach to reasoning to which it's closely related, does not explain how we come to compute human intuitions as recorded in psychological experiments. What it does is provide a mathematical description of which arguments objectively hold based on a common sense that deals with uncertainties and that does not reason computationally with certainties.

As visual and other perception starts with uncertainties, the common sense part of human reasoning starts with uncertainties as well. In both cases, the Edgingtonian conditional is the device by which uncertain inputs get coalesced into regularities useful for increasing individual fitness. In line with this thesis, there is then at this level no need – and rather an inconvenience – to talk about (bounded) rationality or heuristics. Human commonsense reasoning cannot in any meaningful way be judged against strategies to compute in some symbolic representational dimension in one's head. As it was put in Wittgenstein (1998): *"There is no religious denomination in which so much sin has been committed*

through the misuse of metaphorical expressions as in mathematics." Common sense, rather than being rationality or mathematics embodied, is the fallible perception of an outside world that admits of mathematical description. It is to be judged based on its ability to abstract a certain set of useful regularities from reality. The latter will be treated in more detail in chapter 3.

Second, common sense so construed inherently leads to a directional, goal directed, way of being in the world. Any cue can give rise to a certain "inference" – this is the sense of common sense as is explicitly contained in the Bayesian construal of the conditional. Such inferences are by design amodal; the ontological status of A, in "If A, B", implies nothing about the ontological status of B. They are therefore also, without complexity, domain general. This also means that conditionals as construed here suggest a basic mechanism of active perception. Given that "Given A, B" means what it is here taken to mean, one does not need an advanced reasoning faculty to derive a basic "intentional stance" coupled to the basic notion of feedback and basic survival strategy of trial and error (see Dennett (1994) on his "Popperian" creatures and Jo Bervoets (2005b) on artificial life).

To develop the parallel with Gibson (2001) even further, without attempting to explicitate it fully right now, conditionals allow to "*capture*", or "*pick up*" or "*resonate*" with higher order invariants without the need to pass through computations, or representations or any other argument or inference going to *"the desirable via the possible"* (see D. Marr (2001)). As we will see later, there's no need to construe cognition and perception as being composed of 2 different types of stuff. Rather, they can be construed as being composed of a single type of stuff – the historical dichotomy being justified only by what they are "about", i.e. visual/physical with respect to social/linguistic perception. I will try to argue later how complex processes such as selective attention and consciousness can be adequately captured making reference only to the internal construal of the type of conditional here discussed (see chapter 6).

Third and final, with probability – more correctly put: probabilistic structures – being suggested to be at the basis of all cognition one is startled by the apparent purity of some of the pure reason humans produce. The paradox can maybe be paraphrased as follows: "how can certainty be our social currency, if the most money can buy is lowering your degree of uncertainty". Phrasing it in terms of the present study: how can probabilistically construed conditionals lead us to bodies of scientific knowledge whose verification or falsification is mainly based on the mechanisms of pure – mainly deductive – logic?

A common genetic basis and a common living environment are sufficient conditions for individuals of a certain species to arrive at similar conditional beliefs. If such creatures additionally develop adaptive social behavior – e.g. imitation – linguistic exchange of these conditional beliefs of conspecifics is an effective and cheap short-cut to building those beliefs from the bottom up. From a certain point on, the spreading of certain conditional beliefs – quite akin to an epidemiology of representations as described in Sperber (1996) but with "viral" conditionals – is hampered by the inherent ambiguity and fuzziness of the "natural" conditionals formed under extra-linguistic pressure (by implicature only, so to speak). It seems that what is needed to bridge the gap to a real linguistic culture is to find a certain conventional, pure or formal treatment of these conditionals: unambiguous and independent of specific background knowledge. But, as argued in chapter 4, such conventionalization and purification can be conceived as an external pressure on communication or social interaction, i.e. it can be conceived as a fact about the use of language without needing to posit additional psychological faculties or brain organs. Consistent social use of conventional unambiguous version is enough to build pure reason on, the psychological capacity of common sense (using practical conditionals) suffices to account for the bridge function[14].

But let us now turn to the substance of the two last points by building a foundation for our house in the next chapter and by looking for a convenient roof for it in chapter 4.

[14] Small wonder then that the progress of analytical philosophy is from clarifying the conventional and therefore more accessible interpretation of the conditional to the clarifying of the practical interpretation of the conditional. One could say even that the practical conditional "affords" the material conditional, with affordance an ecological rather than a purely psychological concept.

Chapter 3 – The "common" in common sense:

"What do we need probability for? We want probability as a guide in life.
For it to function as a guide, we require it to be objective rather than subjective."

H. Kyburg & Choh Man Teng, "Uncertain inference", 2001.

Given our house appears to be constructed from world knowledge under the form of conditional beliefs as construed above, it is tempting to account for human reasoning in a subjective Bayesian framework of probability calculus. Specifically, it would be tempting to account for a logic of uncertainty, as per DE1, in line with the principle of probability preservation or supervalid arguments in the sense of E. W. Adams (1975). This temptation corresponds, I think, to an important tradition in cognitive psychology interpreting phenomena under the framework of "hypothesis building" and "hypothesis testing" (see Papathomas (1999) for such work in perception, and Komatsu (1990) on explanation-based conceptual structures) and, generally, all theories from mental models to relevance where something is constructed "at the top" to be later on verified or falsified by "deeper down" functions in the brain.

In line with my basic conjecture, I can't admit of such a top-down account (still functionalist at heart) as it would be tantamount to wanting to build a house without foundations. Indeed, in interpreting "If A, B" one does not exhaust the explanatory burden by taking interpretations of "A" and "B" for granted. It is particularly uncomfortable here if, through the back door, we'd need to admit advanced calculations to come up with the assumed meanings of "A" and "B" in the first place.

The present chapter does not admit being read separately (but see Jo Bervoets (2005c) for a stand-alone development of it). A first section 3.1 is concerned with providing objective unconditional foundation of the A's and B's required to state basic "starter"-conditionals. It follows the relatively recent work on uncertain inference by H. Kyburg & Choh Man Teng (2001), chapters 9 to 12 thereof – henceforth HK. In section 3.2 I will use this theory to review 2 basic strands of psychological experimentation. Section 3.2.1 considers research on probabilities and decision-making whilst section 3.2.2 considers the field of categorization and induction. Finally in this chapter, section 3.3 clarifies how, on my view, this account of uncertain inferences, if properly applied, explains the emergence of a "purely" natural language that is common to like creatures raised in like environments. Working up to chapter 4, this section will also indicate how the latter language, when combined with the conditionals of chapter 2, allows emergence of a scientific "pure" language, capturing in explicit computations the inductive logic as per HK.

3.1. Kyburg's evidence for evidential probabilities:

I will not be able here to do justice to all aspects of the theory presented in HK. I won't go into the logico-mathematical detail of the theory, nor into the related semantics he provides. They are, in line with my present thesis the least likely to be cognitively relevant. The underlying regularities of the world – the implicit semantics, if you will – are captured the syntactical features that are in many ways given directly to the senses. As Kyburg himself puts it on p. 286, his approach is an alternative to:

"One common response to these uncertainties has been to retain the structure of classical deductive logic, importing into the premise set whatever is needed. If what is imported is plainly questionable, it is called an "assumption," the idea being that no one can be held responsible for making an assumption so long as it is plainly labeled as such. But an unwarranted assumption is, in our view, no less unwarranted for being admitted to.

Another common response is to insist that uncertain inference be replaced by the manipulation of probabilities in accord with the classical probability calculus. We have suggested some drawbacks to this approach in earlier chapters. The most serious, from our point of view, is the fact that it depends on probabilities that reflect no more than subjective assessment or stipulated arbitrary measures on a language. Such probabilities undermine the objectivity, both logical and empirical, we take to underlie justified inference."

The alternative provided is that of evidential probability. It ties together an objective frequentist view on probabilities – as empirical frequencies observed in the world – with the inferred probability of this

or that specific event – the type of probability we need as a guide to life – based on *the evidence that is provided for in the observed frequencies.* The theory therefore purports to go beyond strict frequentist views in assessing probabilities of specific events without falling into a strict subjective Bayesian view of probability. In other words, it purports to provide precisely what's presently, for my account, needed as the foundational element.

The approach taken in HK has a number of distinctive features, required to bridge the gap from the frequentist reading of probabilities to probabilities of specific events (or vice versa), as follows[15]:

- The deciding issue in uncertain inference will be the choice of a suitable reference class. It is only when considering a certain event as one of a series of *relevant* like events that one can associate a non-trivial evidential probability to the event.

- Probabilities cannot be known precisely. Using the traditional approach of certain point values (real or rational numbers) to label uncertainty does not capture uncertainty at all. Therefore, it is proposed to use interval-valued probabilities, expressing the approximate knowledge that e.g. coins land heads *about* half the time.

- This means a rather radical shift in what is deemed to be a vacuous probability statement – in case no evidence, no suitable reference class emerges. This is no longer a probability of ½, but rather the interval [0,1]. As. Kyburg notes this aggravates the problem of the reference class; any reference class can be relevant to any future event, as any reference class could potentially reduce the interval [0,1] somewhat.

These basic features require the development of a novel, principled – or logically prescriptive –, way of dealing with knowledge derived from experience. As per HK, p. 201:

"(..), we need an inductive or nonmonotonic logic: a system of rules spelling out the conditions under which uncertain statements can be accepted as a basis for probabilities and the conditions under which those statements must cease to be accepted, and replaced by other statements."

This view of nonmonotonic logic is quite distinct from the canonical views on nonmonotonic logics (see H. Rott (2001)). Nevertheless, I need to make my, by now usual, warning against a direct cognitive interpretation of this theory. I do not wish to contend that we calculate according to this specific set of rules any more as I'd wish to contend that we argue according to traditional logical or modern nonmonotonic rules. My aim here merely is to argue that the type of theory developed by H. Kyburg, and further summarized below, is the proper and normative way to deal with uncertainty and that it therefore should provide an ideal basis for describing ways how humans – and potentially many other higher mammals – deal psychologically with uncertainty.

The psychological relevance of such a theory can be phrased simply as per HK, p. 201:

"What do we need probability for? We want probability as a guide in life. For it to function as a guide, we require it to be objective rather than subjective. Two agents, sharing the same evidence – that is, agreeing on the objective facts about the world – should assign the same probabilities. (We do not assume that they assign the same utilities, and thus we do not assume they make the same decisions)."

We need human reasoning to deal uniformly, in principle, with questions of uncertainty. Without such an assumption we would never be able to make sense of human common sense reasoning at all. That we do this primarily via a syntactic account shouldn't be surprising – the syntactic account captures all that is potentially objectively relevant to dealing with uncertainty. As Kyburg notes, introducing terms like "projectibility" – as done by Pollock, and as closely mirrored sometimes in a psycholigistic use of terms as "relevance" (as opposed to the current objective use of it) – begs the question if we are not able to make it immediately explicit how projectibility or relevance is picked out syntactically.

Skipping the logico-mathematical formulation of the theory formulated in chapters 9 (syntax) and 10 (semantics), it is still possible, in my view, to informally convey the basic principled reasons selecting a specific reference class out of the numerous candidate reference classes. What we need to convey –

[15] I follow here the introductory comments of HK, see his chapter 9.1. I refrain from using all quotation marks for ease of reference.

see HK p. 211 – is how we are to select (make precise) the relevant properties of the target that will allow us to establish some well founded evidential probability based on the frequencies observed in the various candidate reference classes. We need to find, see HK p.212:

"(..) grounds for ignoring information (..). We have already mentioned that some information may obscure the import of other information. We need principles by which we may judiciously disregard some of our evidence."

H. Kyburg gives three such grounds, each expressible syntactically in a rather straightforward way as rules by which some evidence may *sharpen* other potential evidence. Such *sharpening* will obviate the need to take into account any evidence that can be sharpened by other evidence as well as the need to rely on things, assumptions that are not presently part of our evidence. The end result of applying such sharpening is an evidential probability, on the basis of the current evidence, of a certain proposition. If some future evidence is imported this may, nonmonotonically, lead to another evidential probability. The distinct forms of sharpening are:

- Sharpening by precision: any knowledge based on a statistically more significant sample will sharpen any other knowledge insofar as the evidential probability interval of the former is fully comprised in that of the latter,

- Sharpening by specificity: if the evidential probability interval suggested by a more specific reference class conflicts with a less specific reference class (i.e. the former interval is neither comprised in nor comprises the latter) the more specific knowledge rules.

- Sharpening by richness: if the evidential probability interval provided by the "more informed" knowledge conflicts with knowledge of a purely marginal kind, the more informed, "richer" knowledge takes precedence. Referring to HK, p.217, it may be clearer if explicitly stated as *"It is this condition that allows Bayesian arguments to take precedence over arguments based on marginal frequencies, when they are appropriate, that is, when the resulting intervals conflict with the intervals that emerge from the marginal distributions."* Whether or not the conflict arises depends on the extent to which a Bayesian argument itself is supported, i.e. how narrow an evidential probability interval it yields.

Obviously, computing which reference class – more accurate: which statement of evidential probability corresponding to a specific reference class – takes final precedence requires settling a computation, or procedure, see section 9.7 in HK. In keeping with the scope of the present work, the above informal statement of the diverse ways of sharpening suffices, more so since I do not wish to suggest a close analogy between the computation proposed and human commonsense reasoning. I only wish to suggest that commonsense reasoning can be explained qualitatively in line with these principles of sharpening.

That this is not prima facie implausible is readily demonstrated by quickly looking at some hallmarks of the reasoning literature and some corresponding basic tenets of nonmonotonic logics. Specificity is taken to be a key aspect of most nonmonotonic logics preferring "by default" more specific knowledge to more general knowledge. The idea is in fact so entrenched that related ideas such as "default inheritance hierarchies" are being used in work on memory (Baddeley (1999)), on conceptualization (Komatsu (1992)) and up to linguistics (Sag (2003)). The prevalence of specificity is such that it can be traced to important modern developments in computer sciences (type hierarchies in OO programming).

But this focus on specificity in nonmonotonic reasoning is already falsified by the already cited work of Ford, Ford (2004). She takes aim – with great precision – at treating specificity as only nonmonotonic adagio discernable in common sense reasoning. We have outlined above two potential basic alternative maxims (sharpenings) at work in nonmonotonic reasoning.

As Ford's work is concerned primarily with reasoning about "blank" predicates (i.e. where there is no prior knowledge that bears conditionally on the predicates interpreted as part of the "real" or "known" world), her research should show specifically the importance of "precision". That this is so can be seen from the effects of adding size info to the problem and is also immediately apparent from the protocols of participant's responses she gives. But also sharpening by richness can come into it indirectly. Indeed, see chapter 2, there can be conditional acceptance of arguments as more or less valid (which cannot be blanked out by choice of certain blank predicates). Again, a close view on her results and the

participant's protocols reveals just this influence. More detailed illustration of the interplay between the different principled forms of sharpening will be given below in section 3.2.

To drive home the point of objectivity that is crucial to Kyburg's theory – as it is for my own, see the next section – let me use another quote of HK, p. 226:

"(..) In contrast, there is nothing subjective in evidential probability except that the relevant evidence may vary from individual to individual – from subject to subject. Being confident that something is so does not make it probable.

There is no dependence either on precise "degrees of belief" or on a logical measure defined over the sentences of a language. This last point deserves special emphasis. On the view developed in this chapter no prior distributions are defined or needed. Probability is dependent on the language we speak, through its dependence on the classes of reference and target terms. But, first, this choice of a language does not commit one to any probabilities at all: with no empirical background knowledge, the a priori probability of any empirical statement is [0, 1]. And, second, we will argue later that our interests and our background knowledge will lead us, not to an "ideal" language, but to the grounded choice between two suggested languages. We will argue that this entire procedure, from determination of probabilities to the choice between two languages, is objective in the sense that any two rational agents sharing the same evidence will be bound by the same interval conditions."

This extract puts the right stress on the key points of objectivity and importance of choice of language over and above the calculus of measures of belief within a language. Nevertheless, it shows a number of issues that remain in order to apply the theory, specifically if applied to common sense reasoning.

First, it risks conflating "rational agent" with the workings of the human brain. In fact, as with any normative theory[16], one should take care to divide the theory from its embodiment. Whilst the former can theoretically prove how an ideal rational agent "ought" to reason, the latter is limited practically – by the specific constraints posed by the mechanisms implementing it – to converging to it, in such a way that the theory can be used as a description of the practice. We need to posit a human, all too human implementation of the theory. One that avoids talking as if consciousness, or reasoning in a pure critical way, is involved (see chapters 4 and 6 below). One that avoids "grounded choices between two suggested languages", at least as far as asserting such choices of psychological creatures as individuals.

Second, it can be taken to promise more than it delivers, or can deliver. Whilst the procedure may be clear and objective, the actual application of it may be a matter of considerable variability. There is no guarantee here – as there is no guarantee in other theories of pure reason – that the actual constraints of carrying out the procedure will not have considerable impact on the intermediate results (or, to put it more correctly, on the intermediate choices to be made). The procedural and declarative meanings of the theory may diverge enormously (as is the unfortunate case in SW development). Due consideration of the initial conditions of the process in actual human beings as well as in actual scientific progress should temper our enthusiasm as far as the attainability of "objective" is concerned. When Kyburg attempts to refute on p. 271 the below aphorism by C. I. Lewis:

"there can be no probability without certainty."

I think he errs on the side of overestimating the extent to which his theory can account for all elements of common sense reason, let alone of pure reason. It will be the subject of section 3.3 and chapter 4 to clarify which elements outside of his theory – a priori elements so to speak – are required to set up a system of agents that can be conceived of as carrying out the procedure as stated above *without* principled bounds of the type psychologically embodied in humans. But as the latter is the actual subject of this chapter, let me try to pinpoint – still following the basic framework provided by H. Kyburg – how the above issues reflect on psychological abilities related to common sense reasoning.

The central practical issue in any psychological application of the theory of evidential probability is the well-known problem of "selection". In fact, the issue of ("intuitive" as against "conscious") selective attention is in my view coextensive with the problem of selection of "relevant" background knowledge. The cursory treatment of it in HK p. 225, does not exhaust the issue. As he says himself on p. 237:

[16] That's the essential truth provided by the work on bounded rationality as per Newell (see Todd & Gigerenzer (2003), Kahneman (2003) and many more).

"But the problem of the reference class is central – and insuperable, because, as we shall see, it has no direct solution.

For exactly the reasons adumbrated in the last chapter, we don't want to be concerned with all supports sets (..), but only those that are relevant. (..)"

But the problem is that at first glance all reference classes *are* (at least potentially) relevant. How would common sense avoid a problem very similar to the frame problem on account of this theory? True, some reference classes are deemed irrelevant on account of the theory, but only after considering them in the context of the rest of the potentially relevant background knowledge. It seems that, after all, we are stuck in exactly the same type of trap that we reached, albeit more rapidly, in the case of deductive logic. The solution, of course, simply lies in not expecting psychological agents to deduce the best reference class based on all background knowledge. Psychological agents have to make do with overt or surface cues that can be relied upon to "steer" the selection of the immediately apparent reference classes, after which some process akin in behavior to the above theoretical procedure can be followed to come up with the best evidential probability (and, equivalently, the best reference class).

To take an example from HK, p. 249:

"The fact that ducks are smaller than geese should not, absent knowledge that size is relevant, undermine analogical arguments based on their properties as water birds."

But, although this is impeccable as far as intuition goes, the absence of knowledge that size is relevant can only be had by searching all knowledge. And, similar to negation-by-failure in theorem provers (see footnote 1), the fact that this knowledge is absent does not mean that it should not be deemed relevant to search for this knowledge – a type of behavior that is certainly characteristic of human common sense as evidenced by the commonplace "don't judge a book by its cover".

What I suggest is that the three sharpenings, as well as their combination, are characteristic in a very basic way for human mechanisms in both what is traditionally called cognition and perception. Where, initially, they are used passively they do allow, with the acquisition of conditionals as above construed, a progressively more directed, or if you will: intentional, stance (Dennett (1994)). Used passively, the knowledge that can be attained through this mechanism is limited to what can be ascertained from direct perception – it concerns the type of intuitions uncovered by work in categorization and judgment reviewed below. The range of knowledge that can be so constructed is heavily constrained as the only elements that'll steer the process of selection, or selective attention, are elements that are directly given in sense perception.

But, as treated in chapter 4, once this attention can be actively, consciously, steered in a critical way the range of knowledge that can be constructed admits of no other bounds than those that are implicit in the type of terminology or notation that we have conventionalized. Although behavior is so drastically different between these two uses of sharpening mechanisms, the mechanisms themselves need not be very different at all. In fact, to jump a bit ahead: the only thing that needs to be drastically different is the environment in which these mechanisms are used.

I restrict myself in section 3.2 to the passive uses of these sharpening mechanisms, showing existing experimental items admit interpretation using HK as a descriptive theory of intuitive human behavior. I will come back to the introduction of an active use of them in section 3.3 working up to chapter 4.

3.2. Empirical evidence for evidential probabilities:

I am unaware of any work in experimental psychology considering a framework as presented in HK. Nevertheless, it should be possible to use it in a rather straightforward way to the psychologistic work explicitly concerned with assessing the value of principles put forward by nonmonotonic logics. I will however for reasons of succinctness, limit myself to other strands of research. First, in section 3.2.1, I deal briefly with the mainstream contention that people do not fare well with the probability calculus (Samuels et al. (1999), Kahneman (2003). Second, in section 3.2.2, I attempt to show how the above theory can serve to explain the empirical findings in categorization and induction – following the work

of Osherson et al. (1990) – henceforth DO – with as specific aim to account for the recent findings of Medin et al. (2003) – henceforth DM – and the implications of HK for "relevance" frameworks.

3.2.1. Probable and improbable probability calculus:

It is a well established empirical fact – Slovic et al. (1988) – that humans have difficulties, and in many cases insurmountable difficulties, to respond according the most rudimentary probability calculus, even when faced with some extremely simple problems such as the, by now canonical, Linda problem[17]. This being as it may, the above view does not admit of any straightforward application of a probability calculus in uncertain situations, i.e. situations where one cannot be certain that events are characterized by specific point probabilities. To quote H. Kyburg quoting Reichenback as per p. 200 of HK:

" "(..) we may have reliable statistics for a reference class C, whereas we have insufficient statistics for the reference class A.C. The calculus of probability cannot help in such a case because the probabilities P(A, B) and P(C, B) do not determine the probability P(A.C, B)"[Reichenbach, 1949, p. 375]. The best the logician can do is to recommend gathering more data."

Or, consider, on page 220 of HK):

"Complementation holds in the obvious way for evidential probability, but that is the only relation among evidential probabilities that can be obtained without more explicit consideration of the contents of «symbol for» the background knowledge."

This suggests that in real-life situations (as opposed to "pure reason" situations calling for applying the probability calculus) when faced with uncertainties it is – on this theory – more often than not indicated to look at the merit of the problem "in context" instead of at a calculus based on assumed probabilities. Given that it is the purpose of experimental psychology to gauge intuitive responses of participants, one can only expect those responses to be in line with the description offered by a passive use – as defined above – of the principles of sharpening according to HK. A more considered response of participants well involve them justifying their answers explicitly by using external tools such as the probability calculus but, as pointed out in Kahneman (2003), such more considered responses tell us more about the (social) background knowledge of the participant than about her psychological functioning.

If we then can establish that the phenomena empirically observed within this range of research are quite lined up with what we would expect from a passive use of Kyburg's theory, we may be entitled to drop various charges of irrationality leveled against humans on this topic. Instead we may offer the passive use of Kyburg's theory as the commonsensical approach to uncertain reasoning where (HK, p. 231):

"(..) we are not concerned with the characterization of some form of truth preserving inference. (Note that inference within the framework of the probability calculus – from P(A) = p and P(B) = q to infer P (A&B) is smaller than or equal to p+q – <u>is</u> truth preserving."

Take, for instance, the cab problem (see Slovic (1988), p. 682), which is taken as a paradigmatic example of the base-rate fallacy (or base-rate neglect). Kyburg [personal communication] says in this regard:

"(..) One thing that is striking is that what Kahneman and Tversky call common mistakes need not be mistakes; e.g. there are natural prior probabilities that make the popular answer to the taxicab problem the right one, (..)"

The issue here obviously is one of arbitrating between different types of evidence. If, for whatever correct or mistaken reason, a subject has stronger evidence for using the simple probabilities instead of the ones conditionalized on the proportion of cabs then the latter will be sharpened away by precision.

[17] I will be mostly referring to the overview article by Slovic (1988). The Linda problem is reviewed there. It basically consists in describing a person "Linda" as having obviously feminist characteristics. The participants are then asked to rank statements by increasing likelihood. The empirical fact, labeled conjunction error, is that most people rank "Linda is a feminist and a bank teller" as more likely than "Linda is a bank teller" in obvious contradiction to the conjunction rule of probabilities (a conjunction of "A" with "B" cannot be more likely, normatively, than "A" taken in isolation.

Such a state of affairs may for instance apply if the subject assumes the simple probabilities to be the result of measurement within the target environment. Quoting Slovic, p. 683:

"Bar-Hillel (1980) offered the most encompassing explanation of the base-rate fallacy proposing that: "… subjects ignore base rate information when they do, because they feel that it should be ignored – put plainly, because the base rates seem to them irrelevant to the judgment that they are making." (p. 216)"

In the framework presented here this translates quite straightforwardly to the fact that probabilities that are obtained by conditionalizing on base rates is sharpened away by what is judged to be a more precise probability based on direct measurement in relevant circumstances. I do not wish to be interpreted that this account of why the base rate is neglected is itself a reason that it ought to be disregarded. Whether or not it ought to be neglected depends on the evidence of the case and the issue in interpreting these experiments lies in the fact that they are underdetermined, i.e. that the information explicitly given does not suffice to make a final judgment. In such a case common sense would have it – as would Kyburg's theory – that explicitly given information needs to be extended with whatever background knowledge may be indicated to be relevant by the types of cover stories used in the experiments.

Such a situation obviously "infects" the experimental item with beliefs that, in the sense also indicated by Ford (2004), allows participants to short cut the need for probability calculus. In this case it infects it with the belief that simple probabilities directly given are considered to be measured in conditions that are the ones applicable to the judgment. The experimental result maybe tells us something real on a psychological mechanism underlying such judgments: humans prefer to rely on probabilities given "at the surface" because they are deemed to be more "precise", unless there is info the contrary.

Obviously, I neither wish to be interpreted as saying that, when the experimental item would be fully determined – and consequently the Bayesian calculus required –, people would fair well. They would fail – any examiner in the field of probability and statistics can tell you – because they would not have the ability to carry out such calculations (certainly not without the benefit of a calculator). I guess that most experimental subjects in such a case (e.g. when they have reasonable reason to doubt that the direct probabilities were measured in conditions that reflect the actual situation) would opt for "don't know" or for guessing probabilities somewhat lower than the direct ones without calculating (and many would be frustrated for not getting a calculator when the case obviously indicates its need, I suppose).

Much has been made of the supposed ease with which subjects solve "frequentist" representations of a similar problem, see Todd & Gigerenzer (2003). But the bare fact is that such representations provide more information to the experimental subject, allowing him to assess the different pieces of evidence in a more "logical" way, independent of any far-reaching philosophical considerations. As is clear by now I believe that the experimental results can be explained by a relatively unsophisticated process in which the relevant items of background knowledge are "brought to mind", after which sharpening is used to judge between them. Conditional beliefs may well come into this indirectly, not via calculation relative to a presented problem but rather via retrieval of already established conditional statements relevant to the presented problem. This is apparent from the work on debiasing, see Slovic p. 686:

"In debiasing research, an important distinction is between producing better judgments and producing better judges. (..), training or looking for contradictory reasons offers some hope of improving the judgment process and being broadly useful."

In other words, debiasing is a question of enhancing the background knowledge of subjects prior to the judgment. Absent such better knowledge we can excuse people to rely on "the representative heuristic" for the simple reason that it is not, in essence, a heuristic. Absent "richer" or "more specific" reference classes, representativity (or similarity) rules by "precision". Asking people to come up with Bayesian calculations on the spot and without tools for carrying out such calculations is, as I contend, simply asking too much of them, much more in any case than common sense could either justify or cater for. But, fortunately enough, we can stimulate the background knowledge that in some situations relying on one's background knowledge doesn't cut it. In such cases, if the option is left open experimentally, one should not be surprised to see subjects naturally rallying towards paper and pencil.

Although it certainly is interesting to try to account for more results of the heuristics/biases literature on the basis of our three mechanisms of sharpening (e.g. subcertainty in Prospect Theory, Kahneman &

Tversky (1979)), I cannot here go in any detail in them. I will conclude with a cursory treatment of the famous Linda problem[18] as it is a basic challenge to the correctness of common sense intuitions, even when accepting all of the caveats I have posed higher.

The Linda problem is most obviously underdetermined. It is underdetermined to the extent that it wears its putative Gricean implicatures on its sleeve. Indeed, in terms of evidential probability one can readily establish – via extensional "precision" – that the evidential probability of Linda being a feminist is high (in fact, quasi-certain for any Westerner). Nevertheless, one cannot establish anything on the evidential probability of Linda being a bank teller. Donovan & Epstein (1997) clearly interprets that probability to be very low (see Vignette A p. 18) but personally I don't see how subjects are supposed to come to that conclusion.

To me, evidential probability is narrow and high for Linda being a feminist because she is defined as such whilst evidential probability for Linda being a bank teller is the vacuous knowledge of [0, 1]. This being the case – and referring to the quote of Reichenbach stated higher – there is no real arbitration by way of the options "feminist" and "bank teller and feminist". Both have no support beyond the vacuous support and retention of either one seems merely a matter of preference. One can only conclude, in my view, from the Linda problem that people prefer to retain the more informative option - and I suppose that from the point of view of common sense much can be said in favor of them preferring it.

Notwithstanding my reservations with respect to the symbolic interpretation of the Linda problem by Donovan & Epstein (1997), I do believe their general conclusion is in line with the present framework. Quoting from the abstract:

"To determine whether the high rate of conjunction errors (CE's) to the notorious Linda problem can be explained by the violation of implicit conversational rules (..) participants were given completely disclosing information. Although this procedure, directed toward a rational mode of information processing, reduced CE's, a majority of participants continued to make CE's. A graded series of problems designed to activate latent, intuitive knowledge (..) additionally reduced CE's. (..) In certain situations the outcome of the experiential-intuitive mode is more compelling than that of the rational-analytical mode, even when the latter is equally accessible."

Paraphrasing at my own risk, the errors are reduced whenever the situation portrayed is such that there is a natural reference class suggested by the problem itself. In all other cases, where some computation would be required to make sense of underdetermined or otherwise awkward data, subjects tend to have a greater preference for the conjunction because it is the option for which they have some knowledge – i.e. some reference class. Given that the methodology of these experiments precludes subjects from the option of analyzing the problems by means of paper and pencil[19] and consequently does imply subjects to rely on their intuitions, there is no reason for them to second-guess their common sense intuition.

Not enough said on the matter but enough within the scope of this thesis. The bottom-line is that there's no reason to embellish the results obtained in the heuristics/biases tradition. People do judge based on representativity and similarity, as well they should in the absence of any reliable and clearly conflicting information. People don't correct the intuitions they have based on rational analysis when they lack the tools, information or incentive to make such an analysis, as is the case in most prototypical "between subjects" experimental methodologies. Unless heavy training is provided to simulate these tools within working memory, people simply cannot be expected individually to come up with insights that have cost centuries to discover. As a final note, it should be clear from the above that the approach suggested in DE2, following Adams (1975) and carried out in Adams (2002), implying people reason according to probability preservation (and hence implicitly via probability calculation) cannot be correct.

3.2.2. Applying the logic of induction to … induction:

[18] A treatment that is at odds with how H. Kyburg would treat it. He feels this fallacy is to be seen in terms of Gricean conversational implicatures [personal communication], but see further in this section.

[19] It is of some importance to note that, when pressed for reasons or justification as done in Donovan & Epstein (1997), people do tend to select the right principle whilst giving the wrong answer (see, page 11, "the kind of statistical principles endorsed in the Principles Condition). If people are cued to select a principle of pure reason they don't fair too badly. Of course, this is of no great help when it is all but apparent what principles of pure reason would be applicable in a certain case.

Although in experimental psychology categorization and induction is a relatively separate field from that of reasoning in experimental psychology, this shouldn't be so on the present conjecture (for a very similar remark on conditioning and learning research, see KD). As stated in the first line of DM: *"One of the central functions of categorization is to support reasoning."* I do not hesitate to go far beyond that claim – shunning as is my custom any talk of "functions" – and state that categorization provides the first natural proto-language that is a necessary minimum condition for any reasoning to take place (a more thorough treatment is provided in Jo Bervoets (2005c)).

I want to indicate here that the type of reasoning immediately afforded by categories corresponds to the mechanisms of sharpening where sharpening-by-richness provides a Bayesian link with "practical" conditionals as per chapter 2 and their specific role in higher reasoning as per chapter 4. We start from DO, a seminal work on category-based induction. Using natural categories and blank predicates (about which no prior knowledge is assumed to exist), they investigate the argument strength of various arguments, e.g. their example (3) on p. 186:

"Mosquitoes use the neurotransmitter Dihedron
Ants use the neurotransmitter Dihedron

Bees use the neurotransmitter Dihedron

Where the premises are put above the line, the conclusion below it and the argument strength relates to the acceptance of the property in the conclusion category relative to its categorical assertion of it in the premise categories.

A number of phenomena are investigated, each corresponding to a specific form of argument[20]. I focus on, see p. 188:

"Phenomenon 11 (inclusion fallacy). A specific argument can sometimes be made stronger by increasing the generality of its conclusion."

The specific example has subjects judge an argument from robins to birds as stronger than the same type of argument from robins to ostriches. The authors state that such a conclusion is counternormative, which is why it is referred to as a "fallacy". Nevertheless, according to Kyburg's theory here adopted as normative for uncertain reasoning and induction, the fallaciousness of this move is questionable. Whilst robins certainly constitute an appropriate reference class for birds, the same is not as apparent for robins vis á vis ostriches. If this is so – I do not think many people would argue on this issue – the normative result following Kyburg's theory would certainly be to accord the higher and narrower evidential probability to the argument leading to birds. No conflict or inconsistency arises with respect to allowing the property in birds with more confidence than allowing it in ostriches – after all it is well known that the specificity of ostriches sharpens away a number of properties that could be asserted on the basis of the reference class of birds.

But let us try to be more quantitative in our assessments as is readily allowed by the quantitative model in DO, predicting assessments of argument strength based on two variables: similarity and coverage. As per p.189-190 a paraphrase for these two variables goes as follows:

"The theory developed below claims that confirmation varies directly with the following two variables: a = the degree to which the premise categories resemble the conclusion category; and b = the degree to which the premise categories resemble members of the lowest-level category that includes both the premise and conclusion categories."

Hence the name of the model: the similarity-coverage model. My specific contention here is simple: the two variables correspond directly to two of our three mechanisms of sharpening. Coverage sharpens by precision (which is why it is absent in "contextless" experiments reviewed in section 3.2.2). The better the premise categories cover the conclusion category, the higher the authority of the sample and the narrower the probability interval and, finally, the lower the likelihood of precision being sharpened

[20] It is important to note that on top of using blank predicates, the authors explicitly used methodology designed to *"suppress metacognitive strategies"* such as relying on the symmetrical appearance of two arguments to conclude these arguments to have the same strength. The authors therefore were looking for intuitive reasoning explicitly suppressing any learned or conditioned response based on rational or analytical grounds.

away by any other reference class. Similarity sharpens by specificity. The better (the more typical or representative) the premise category resembles the conclusion category, the more likely it is authorative as a reference class for the conclusion category overriding general knowledge if the knowledge based on similarity is strong enough to conflict with coverage. For natural categories, one can assume that the narrowness of resulting candidate evidential probabilities is directly available to all subjects, through experience.

It is clearly outside the scope of this work to consider the detailed mathematical relations between the model proposed by DO and the theory proposed in HK, although doing so would conceivably allow us quite some progress – for instance with respect to the remaining free variable of the model proposed in DO as well as with respect to the detailed mathematical descriptions proposed there for both similarity and coverage (and corresponding implications for neuronal implementation (see Sloman (1993)).

What we do need to cover still is the element of sharpening by richness. The fact that this element is absent in DO is unsurprising. The use of blank predicates is explicitly designed to exclude the influence of "rich" information as per our background knowledge under the form of Edgingtonian conditionals.

However, that such elements may be highly important is the exact subject of the paper of DM, see p. 7:

"In short, typicality and diversity effects are far from common in populations that have considerable knowledge concerning the domain of categories under study.(..) The most salient reason is that often they are instead employing causal and ecological reasoning about the kinds in question."

I contend that the explanation of this phenomenon lies in the mechanism of sharpening by richness[21]. As people gain a certain expertise (become better judges) they gain background knowledge of the form expressible in conditionals (see example p. 10 on diseases and trees assessed by tree experts *"If one gets it, they all get it"*). In line with Kyburg's theory, such knowledge – for better AND for worse – will sharpen away other knowledge based on similarity-coverage (specificity-precision or typicality-diversity), adding the type of flexibility demanded by the authors to account for the inductive effects that cannot be accounted for by the similarity-coverage model as per DO.

Whilst DM proposes to account for the fact here explained by richness within a Sperberian framework of relevance – after reviewing schemes based on hypothesis-testing and Bayesian reasoning – I feel that this is somewhat like putting horse behind carriage. As already hinted at various times above, what we need is an explanation of why certain things are deemed relevant rather than an account of relevance as influencing reasoning. Obviously, relevance is in this and other frameworks further grounded on items such as "cognitive effects" but doing so only pushes down the computational analogies to lower levels.

I believe such actions are unnecessary and confusing. Assuming background knowledge is acquired by forming of conditionals (and ultimately by a richer language – see below) then surface structures of arguments (within the type of language allowing certain conditionals) suffices to pick out the various pieces of knowledge that are relevant. There is obviously then no guarantee that they're objectively relevant other than the fact that the picking out (the language) is formed through a process that deals in an as objective as possible way with the objective facts available. Then again, there's no uniformity of common sense or intuition across populations on specific elements. Common sense is never better as an environment in which it is derived. Improving common sense only takes place after critical scientific inspection, developed for exactly this purpose, has improved the social knowledge on which it is based.

If, as is very plausible and completely unmagical, the various pieces of evidence are picked out by the surface structure of the argument (or situation) then the decision between a selected few of these pieces can be made according to the normative theory above. In certain cases there is little or no knowledge retrieved to arbitrate between, the case is underdetermined. In other cases arbitration between selected pieces of evidence is not yielding a clear winner. In either of these cases, the evidential probability out of the process will be having a wide interval, maybe even close to vacuous knowledge. In such cases,

[21] This is obviously not all that can be said about the matter. When DM refers to the Itza Maya there also may be an influence of a too broad assignment of the label "natural kind" to categories such as mammals. If, in certain cultures, no such kind is broadly established, reasoning by coverage (diversity in the terms of that paper, precision in our terminology), will be less available a mechanism than in cultures heavily influenced by taxonomic categories having attained natural kind status. Language is not fully independent of culture (although it is in some minimal way so, see section 3.3).

one will have to attend actively to other aspects of surrounding structures or actively look around for more external evidence, until such time the evidential probability is narrow enough to take action (e.g. activate a conditional script or simulator).

Indeed, the local context of any perceptual or cognitive input suffices to pick out, via correspondingly cued conditionals, the relevant information to that context. The relevance of the information so picked out does not need to be guaranteed by powerful computations over symbolic elements as its relevance will be guaranteed by the survival and entrenchment of the conditional so picked out[22]. Whilst the conditional itself can be seen as giving the "sense" of knowledge, the background whereby the conditional belief was built up can be considered as the "common" part of knowledge. The former is useful because it can be picked out based on a local context whereas the second is required because it integrates the broadest possible context of experience.

3.3. Sharpening the "common" in common sense:

What I have hoped to show is that there is a principled and objective foundation to A, and to B in "If A, B". To quote HK (*) a last time (see p. 271):

"(..) How can we avoid infinite regress on the one hand and subjectivity on the other?

The answer is that at some level we have the same evidence, or that we are we willing to share evidence. With regard to the former, we are all born in the same world, and have been subject to very similar streams of experience. This is a matter of foundational and philosophical interest, but not particularly germane to the justification of particular inductive inferences. The latter consideration, the willingness to share each other's evidence, on the other hand is central to the scientific enterprise. It is this that makes matters of integrity and objectivity so crucial in the social fabric of science."

Even the best painters will, in trying to capture the essence of something, use a brush too wide to allow capturing important nuances. H. Kyburg here states a truth but misleads us as to another important one. The two ways ensuring objectivity he refers to above – having and sharing evidence – are maybe alike as far as the mechanisms underlying uncertain inference. They are however fundamentally different in a way extremely germane to the present discussion.

Having the same evidence – experiencing the same environment – provides, together with sharpening mechanisms discussed above, a cognitive basis for establishing a similar language of categories, which allows sharing, inter-subjectively rather than objectively, knowledge according to the passively shared environment. It allows to extract what's common in certain ways of experiencing common environment rather than allowing to extract what is common in that environment per se. What it affords is common sense rather than pure reason. Very much as proposed in Barsalou (1999) it allows perceptual symbols that are however still constrained by the way in which that perception is structured (the way knowledge is embodied, in his terminology). Although a powerful instrument – provided enough perceptual power feeds it –types of background knowledge, including construction of specific conditional beliefs (and elaborated scripts or schema's derived from such conditional beliefs) is limited to what is given.

Culture can, over time, contribute a large amount of knowledge to what was initially given by nature only – certainly when that culture includes a linguistic tool – but that can at most account for common sense in its basic individual form: the derivation of what is common in different streams of experience and the adoption of certain directions given – or conditional upon – detection of such commonalities. All this will be common to like subjects (where "like" depends on having like cognitive mechanisms, but also on them having like natural and social environments), but only up to the inter-subjectivity as mentioned above. Objectivity is not guaranteed, as should be immediately apparent from the fact that common sense is fluid and diachronically evolving under the influence of newly acquired ways to perceive and act in nature as well as newly acquired conditional beliefs. As per the quote of B. Russell opening this work – common sense in itself will not allow explaining the world.

[22] This is by no means intended as a full account of human cognition. Rather, at the present level of discourse it is intended to be a partial but highly significant account of the cognition of higher animals. The feature of selective attention is certainly not behaviorally restricted to humans. What is, or seems to be, behaviorally restricted to humans is the interplay between selective attention and linguistic items controlling it. See next point in this wrap-up for a first glance of extending this theory to that realm.

What is required for objectivity – pure reason in my terminology – is "sharing" of evidence. As pointed out by HK in Chapter 12, the concept of sharing evidence is minimally constrained by the possibility of having the same evidence. Nevertheless, having the same evidence isn't at all enough for sharing it.

No, sharing the same evidence requires intentionally constructing specific languages in which evidence can be expressed not only in line with our having it but also in line with how the world is. The table of Mendelejev is probably the paradigmatic example. Without the language, or grammar, adopted through adopting the table we would be unable to share evidence of a chemical nature. The issue of certainty – or pure reason, or scientific inference – is not addressed by a theory of uncertain inference as such. C. I. Lewis was right: no ("scientific") probabilities without (prior, conventionalized) certainty. Practical or commonsensical reason cannot bootstrap its way out of uncertainty and inter-subjectivity just like that, extreme empiricism does not fly (although it does walk admirably).

So, whilst we have found a plausible foundation for our house, we have not found a roof. Rain would pour in and our knowledge would be left unsheltered and frequently blown away by strong winds, with no other avail open to us but building it up again (myth by myth, so to speak) until another storm blows us back to the dark ages. We have however still one card up our sleeves, because we may have found, in chapter 2 above, the type of material that could serve for constructing the necessary roof. We found it in the conditional beliefs expressed by the Edgingtonian conditional already connected in this chapter to uncertain inference. As it was our purpose to account for common sense and the over-all human ability for reasoning as separate from any implementation of reasoning modules, we can however not leave it at having the materials only. In the next chapter we therefore explore how such materials can be brought to the good use of providing shelter to our knowledge, a roof for our house.

Chapter 4 – Certainty in uncertainty:

"(..) to replace an iconic system of representation by a noniconic system will be to introduce a new and more powerful extension of the original system, one which can do everything the former system can do and more besides."

H. P. Grice, Studies in the way of words, 1989.

We don't here need a philosophical foundation for either having or sharing the same evidence. "That" we have and can share the same evidence can be accepted for my purpose here as an evident fact, as something that is commonsensically given. Rather, what we're in need of is a plausible account of how – given that fact – we are able to come up with a language of certainty based on the tools of uncertainty sketched above. They certainly do not at first glance encourage us to be very optimistic on our chances. Both foundation and the house itself seem constructed from impure material and put together in a rather crude and approximate way.

There is every reason when taking this at face value to resign us to the fact that our house is naught more than a cave in which we are shackled, seeing merely the shadows of Pure Reality. Not only is such resignation premature, but it is also the idea of Pure Reality – of towering reason – itself that implies the shackling. In this chapter, an attempt is made to provide a reasonable roof to what is, after all, only a modest and reasonable house that needn't penetrate the heavenly skies. We should be content if our roof provides shelter and a simple way to adapt our house to our changing needs. The treatment in this chapter is unfortunately as unambitious as the goal set above. Nevertheless, for clarity we will follow a structure as was also used for the previous 2 chapters.

To this end, in section 4.1 below, I indicate how, in my possibly naïve opinion, the work of H. P. Grice, Grice (1989) – henceforth HPG – provides the necessary material. The material lies in Grice's work on the relation between non-conventional and conventional meaning. Empirical work based on his essays is not abundant. What does exist is mostly concerned with conversational implicatures. Consequently, in section 4.2, I limit myself to some methodological remarks trying to show Gricean implicatures are too narrow a focus to attain sensible results from the over-all work on meaning as per HPG. Finally, in section 4.3, I will take the liberty to assume the roof is constructed such as to make far-reaching, and perhaps farfetched, conjectures on some advanced features of its human inhabitants, (see chapter 6).

4.1. A non-conventional view on Gricean conventionality:

Grice is not particularly famous for his points of view on conventional meaning. When referred to – certainly in a non-philosophical context – almost exclusive reference is made to his conversational implicatures. Such is the case in KD, p. 32 ff, in the work of Donovan & Epstein (1997) as well as, no doubt, in many similar works of psychology. This leads to the implicature that HPG suggests meaning is purely a pragmatic thing and that conventional meaning is not much more than a, sometimes fruitful, idealization. Something equivalent is true for work like as that of D. Wilson (1996) on relevance, often cited to be related to the maxims underlying Gricean implicatures, as essentially "psychologistic".

This interpretation of HPG, and work inspired by it, as belonging to a pragmatic and/or psychologistic corner leads to two characteristic reactions, neither of which is productive:

- Relativizing everything to certain pragmatic effects allows accounting for phenomena in an ever-widening circle of context and content effects. It for instance allows observed anomalies of human reasoning to be accounted for by hitherto unnoticed implicatures (see H. Kyburg [personal communication] on the Linda problem). These hang as a post-hoc Damocles' sword over all empirical research. Furthermore, since implicatures are interpreted as "inferences" of the interlocutor's intentions, this interpretation implies some substantive reasoning module required for such, sometimes, complex inferences. This is obviously at odds with my thesis.

- Those more logically inclined easily dismiss as "psychologistic", "subjective" and, ultimately, "circular" most of the Gricean proposals. Although this dismissal may in certain cases well be

> correct, one throws out the baby with the bathwater resulting in two camps (see chapter 1). On one hand there are pure logicists assuming a faculty of pure deductive reasoning in humans to account for their pure results. On the other, psychologists assuming some faculty of ecological reasoning in humans that stands supreme challenging the normativity of pure traditional logic. As in any good antinomy, neither side will agree on anything but both implicitly work from the same presupposition: a cerebral faculty of reasoning with corresponding representations and functionalist interpretations of intelligence and the mind.

As will be clear from the way I discussed this, all of this is misguided and does not do justice to the real ambition of HPG. The first introductory essay in HPG shows that Grice was concerned with the non-conventional conversational implicatures as a necessary device for rescuing (in a manner of speaking) the conventional core truth-functional meaning from what are some traditional challenges to it based on informal conversational practice of humans. Quoting the opinion of somebody with maybe a bit more credibility on the matter, see for instance DE2, p. 391:

"Grice famously defended the truth-functional account, in his William James lectures, "Logic and Conversation," delivered in 1967 (Grice, 1989). There are many ways of speaking the truth yet misleading your audience, given the standards to which you are expected to conform in conversational exchange, One way is to say something weaker than some other relevant thing you are in a position to say. (..)"

Although, as discussed in chapter 2, his defense fails for the Edgingtonian construal of conditionals, it clearly shows Grice had an interest in not accounting for all phenomena on pragmatic grounds. On the contrary, his invention of various "bracketing devices" and his general remarks in the introductory essay show that Grice wanted to save conventional meaning from the morass of taking into account an ever widening circle of pragmatic contextual elements that could not be reduced to truth-functionality.

I can imagine that on the challenge brought by D. Edgington on his defense of truth-functionality of the conditional he'd be tempted answer that her interpretation may or may not be correct but that anyway a need for a truth-functional account would not be obviated. Without interpretation of the conditional as a material conditional a state-of-affairs could not be communicated with sufficient clarity and precision for it to acquire a core meaning irrespective of a specific context. Without conventional core meaning there is a level of uncertainty precluding us to build extensive (deductive) argument as is required to find the logical entailments that might serve as a basis for falsification of the stated conditional.

I think, in this case, we can have our cake (pure reason) and eat it too (practical common sense). I base this assertion on the following quote of HPG, p. 358:

"We might be well advised to consider more clearly the nature of representation and its connection with meaning, and to do so in the light of three perhaps not implausible suppositions:

1. *That representation by means of verbal formulations is an artificial and noniconic mode of representation.*
2. *That to replace an iconic system of representation by a noniconic system will be to introduce a new and more powerful extension of the original system, one which can do everything the former system can do and more besides.*
3. *That every artificial or noniconic system is founded upon an antecedent natural iconic system.*

On my conjecture, categorization as discussed in chapter 3 and conditionals as discussed in chapter 2 conspire to allow the "antecedent natural iconic system" to evolve. Albeit that I take exception to the talk of representations and symbols if interpreted as brain states, it is at this level of proto-language that "grounded" perceptual symbols – in the sense of Barsalou (1999) – as well as scripts, simulators and frames emerge. The only thing afforded by this proto-language is implicature: uttering something so as to cause like psychological effects in like subjects. At this level it is entirely unimportant whether one "understands" what one utters or whether one can "infer" the intention of what is uttered (the latter against a "computational" interpretation of HPG). What matters is simply that what is uttered has a predictable effect up to some desirable but necessarily uncertain and imprecise level.

The constraints when using such a proto-language are huge. Communication is ultimately limited (see again Barsalou (1999)) to both sides of the communication being alike perceptually and cognitively, as well as, I need to add, having been raised in like environments. That does not mean that communication

at this level is in any way subjective or private (the idea of private forms of language or "I"-languages is, after all, self-defeating), quite the contrary. Communication is limited fully by external facts-of- life.

Our linguistic history has however led to "new and more powerful extensions" under the form of some explicit conventions admitting of critical analytic treatment. The core of these extensions is expressed as the truth-functional account of meaning. By convention certain propositions correspond to states-of-affairs that may or may not obtain. This is not a very magical extension; the conditional as argued for above admits of only the feeblest of organizations whilst a more idealized truth-functional use of it as per the material conditional allows the full logico-mathematical machinery to do its work. Therefore, conventionalizing is a powerful means to construct "an artificial and noniconic mode of representation". It is moreover, as far as conditionals are concerned, something which is naturally captured as the limiting case of the Edgingtonian conditional. It also leads to the fact that you can do "more besides" as is clear in the recursive compositionality inherent in such truth-functional accounts.

Unfortunately (maybe), this conventionality is foreign to our brains as such. We will therefore never be as good as computers in calculating anything. It also means that computers – as currently conceived – will never be any good at exhibiting common sense characteristic of human intelligence in coping with the uncertainties of the world. The creativity, for instance creativity by analogy, typically afforded by the implicatures of common sense cannot be captured by recursivity, common sense is as supreme in its dynamic, diachronic productivity as is recursivity in its static, synchronic productivity.

The challenge that amodal symbols with conventionalized meanings can't be neuronally grounded does not succeed for the simple reason that amodal symbols are not so grounded. Amodal symbols together with the arguments they afford are only socially grounded. The same mechanisms as above, in dealing with the directly perceptual world, can deal with the linguistic world of auditory and written symbols. We are not pure enough individually to make logical progress but nevertheless benefit from progress as obtained via social critical exchanges. Our benefit is obtained through our partaking in social exchange.

We have here, in an admittedly embryonic form, an argument to treat artificial and natural language on quite different terms, not too dissimilar from the proposals of Frege and Tarski. The difficulty is – and I take Grice's program to be concerned with that – that both uses of language are necessarily intertwined in human subjects. An attempt at disentangling these uses is the subject of the next sections/

4.2. Methodological implications of implicatures:

In order to illustrate methodological implications of the above, let's start from the Lottery's Paradox.

One may explain it by implicature: if somebody says that "A will happen if B and C and D and E and …" then we know he implies that A won't happen because we know that if A's happening is dependent on a long list of conjunctively required conditions its likelihood of happening is low. "That" we know it to be the case need not be based on inferring the intention of the other. It may well be based on the fact that we live in an environment where a list of conjunctively required conditions is commonsensically equated to assigning a low (and narrow) probability to A.

However, one may also explain it by reference to deductive logic together with probability preservation as mentioned in DE2, following E. W. Adams (1975). Assuming all the conventions of deductive logic and probability calculus one can calculate that there is a certain loss of probability on the part of A's happening with each additional conjunctive clause in its conditions. It is even possible to additionally refer to an implicature as follows: based on the adoption of such conventions, the intention of the other in referring to the Lottery's Paradox is to imply A's happening is unlikely.

All such explanations will predict the same behavior and that is not a good thing. It requires us to add arbitrary grounds of parsimony and the like (see the maxim of minimum mutilation used in H. Rott (2001), following Quine). The multiplicity of potential explanations is, of course, not coincidental. All explanations ultimately are based on the fact that they all refer to a single objective world. Still, it does make it difficult to interpret direct experimental evidence (including, that is, for my conjecture, but see chapter 6). Nevertheless, in accepting the terminology hinted at in section 4.1, I hope we may at least be able to critically analyze the merits, or lack of them, of the various explanations.

I think that Donovan & Epstein (1997) admirably shows that this is a fruitful exercise. Similarly, the work of Levinson cited by KD, p. 32 ff, may lead to a more principled treatment of interpretation of experimental items (an interpretation which is, as should be clear from the above, a core issue in the "framing" of empirical results). Finally, referring to an even more authorative source, methodological remarks as per Kahneman (2003) clearly indicate that an explicit grasp of the issue isn't required to see an implicit reason for caution. I can, unfortunately, not develop this in any detail so I'll leave it at just giving a final tentatively clarifying example immediately below.

There is a situation in which conventional and non-conventional use of language come apart in a rather dramatic way. It is the situation where somebody uses the locution "I can't imagine why she did that". Surely this does not mean that one can't see "her" having certain reasons for whatever she did. If that would be the case we'd be committed to the view that "she" was something of a miraculous agent. I'd contend that it literally (conventionally) means that the speakers cognitive mechanism differs in a crucial respect to the "she" referred to in the phrase. Knowing, after explanation and possibly even demonstration, why she did that doesn't invalidate still using the locution. The reasons, conventionally expressed, for "her" doing something are quite different from the mechanisms by which such reasons become internalized. Truth-functionally the "not being able to imagine why she did it" refers to entirely different states-of-affairs than the "knowing why she did that".

2.3.3. The stepping-stone theory of mind and meaning:

So, we found a roof for our house perfecting materials already present in the house itself. It turned out to be very lightweight and flexible. Conventionalizing commonsensical meaning allowed establishing a direct objective correspondence by means of truth-functionality. The roof is in fact so lightweight and flexible it allows much more and bigger rooms and continuous extension to more and more levels. The result is that one easily gets lost in it and, more often than not, misunderstands other inhabitants. It does indeed arguably afford more freedom of pure movement to the inhabitants than the latter could, at first glance, attribute to their own rather pitiful practical abilities of common sense. The combined forces of recursive compositionality and creativity by implication outwitted even the most cunning investigators of this house of reason. Most were attracted by the promising songs from one side, neglecting the other, and many assumed to find all answers in a tiny place above the eyes, or high up in the heavens.

Allow me the liberty of a very preliminary attempt at disentangling these forces, showing how, one step at a time, the edifice can keep on growing both sideways, allowing more inhabitants, and up, allowing more knowledge per inhabitant.

It all starts with the accrual of unconditional background knowledge in the form of categories (chapter 3) leading to a natural proto-language based in sense perception. Such language allows the processes of selective attention to take an active, proto-intentional, stance. The environment can be sampled for item B "given" some other item A is perceived. Consequently the fundamental conditional device is formed (chapter 2) allowing to derive "richer" knowledge based on relevant contextual and conditional cues. Communication can take an inter-active form, well beyond the simple unidirectional social strategies of imitation. Conditionals can be shared in conversation where one individual may imply something that can be captured by another like individual in virtue of the fact that they share a same basis for forming such conditionals (chapter 4).

But the significance of conditionals is still largely a matter of individuals, even if knowledge need now no longer be based on individual perception but can be taken from "common" opinions (chapter 3). Individuals have a more powerful mechanism of selective attention whereby items of any kind are coupled to items of any other kind. Complex behavior can ensue and there can be some "sense" to the actions undertaken (chapter 2). That sense increasingly becomes "common" as more knowledge will be derived from social perception instead of from direct perception. Perception is no longer limited to the physical world, but extends to social items such as conditional statements of what ought to happen.

When communication is sufficiently fluent, cultures are formed where abstract symbols are used to get to specific social effects. Individual subjects become submersed as particles of a bigger cultural entity which mediates their survival chances much more than individual abilities do. The simplest thing that could possibly happen happens (chapter 4). Meanings get conventionalized in order to allow for much of the increased precision needed to support complex social behaviors. Selective attention becomes a full-blown central executive (see Baddeley (1999)), no longer bound to consider items within an actual

context or related by existing conditionals. It can work with conventional meanings productively to discover “from the top” potential insights between previously unrelated chunks of knowledge. But it still is, and will remain, limited in capacity, requiring external tools and social interaction to increase its scope of work. The products of the former are fed back socially to constitute common cultural elements assuming they survive the test of sharpening (chapter 3). Social interaction can by creatively forming a new implicature across individuals increase the span of knowledge to new fields (chapter 2). Between working memory and the vast amount of background knowledge an interface develops that is closely similar to the social, linguistic interface between individuals (see “structions” as per Jaynes (1976)).

But the consequences of simple things can be quite complex. Indeed, with conventionalized meanings relying on socially demonstrable truth-conditions, it becomes possible to share evidence and therefore also critically reflect on evidence and the language in which it is expressed. This process of dialogue – as all other process that are perceived and consequently stored in background knowledge – is naturally taken over and emulated in the process of selective attention. Once the social dialogue is internalized, it is perceived as inner dialogue leading to an implicature of individual intentions and personality. A Turing-machine seems to be at work and it seems to be of the kind allowing a conscious personality to run on it. The original individual creativity based on commonsensical associations seems something entirely different, intuitive and subconscious. The open society is born and with it individuality is more and more socially confirmed up to the point that “I think” becomes the only ultimate certainty (see, for a less metaphorical conclusion, chapter 6).

We have not only build a house of reason. To our surprise we got the inhabitants for it in the bargain. But with the inhabitants comes the confusion about the house and the type of persons living in it (see chapter 1).

Chapter 5 – An experimental thought and a thought experiment:

It is far from easy to find connections between different theoretical and empirical elements. It is harder still to allow yourself to be convinced by the apparent power of those connections of a counter-intuitive conjecture (and isn't it intuitively the nature of a conjecture to be counter-intuitive). But then and only then comes the hardest part: proving it (or at least proving something in it can be disproven). I have to apologize straight of the bat that this chapter, being about that, is a very short one indeed. Certainly, I'd hope to be excused on the grounds of insufficient time but, quite apart from that, I feel I need to argue forcefully that, on the grounds of my own conjecture, such hard work cannot and should not be left to a single person. The individual creativity I hope I've shown goes only as far as it goes, and for purposes of this chapter that is not quite far enough.

This being said, I will try to show below some initial ideas on how to make a pudding in which one can find a proof. In section 5.1, I address how one maybe could empirically dissociate intuitive responses to a reasoning task based on individual common sense and more conventional, socially justifiable responses based on the emulation of social practices. Such a dissociation obviously does not, and most probably cannot, demonstrate directly my claim that only the first commonsensical behavior is based on dedicated type of brain circuitry whilst the second is merely parasitic on a general-purpose process of selective attention or working memory (see chapter 4). Using the traditional methods of behavioral experimentation, the latter claim is – as far as I can see after many glances – beyond reach. Therefore, as neither brain imaging nor anthropological are within the scope of this work (nor are they within my skill set), I provide in section 5.2 a thought experiment complementing the proposed dissociation such as to make the "pure reason as parasitic on human impure common sense" idea more plausible[23].

5.1 Dissociating common sense and pure reason:

Above I have tried to account for traditional results in the heuristics-biases tradition (see Evans (1993)), in a way commensurate with the different building blocks contained within my over-all conjecture (see specifically chapter 2 and section 3.2.1). I have suggested throughout the 3 chapters on these distinct building blocks how specific tests of these sub-hypotheses could be made (see also Jo Bervoets (2005c) in relation to section 3.2.2) and other tests should, in my view, be able to be derived rather directly for each of these sub-hypothesis. In this section however, intellectual honesty requires me to provide not a test of specific sub-parts of this thesis but a potential test for the over-all conjecture that humans do not reason, hence do not have a reasoning faculty operating, according to the rules of pure reason, whether syntactic or semantic.

The approach indicated for the latter purpose is to show dissociation in participant responses to the standard tasks of the heuristics/biases literature leading to "fallacious responses" (see, for an overview, Samuels et al (1999), Stanovich & West (2000) and Kahneman (2003)). The big practical advantage of this approach is that a baseline of relevant empirical results is available on individual commonsensical or intuitive response behavior of participants. This baseline is furthermore interesting since it has been the subject of intensive scientific critique and since, as shown higher, it admits being explained in terms of the sub-hypothesis provided in chapters 2 to 4. Last but not least, this specific field has a tradition of wide criticism both as far as methodology is concerned (see a.o. Kahneman (2003), Donovan & Epstein (1997)), suppressing "meta-cognitive" strategies, and as far as interpreting the results are concerned (see Stanovich & West (2000) and Sperber et al. (2002), Samuels (1999), just to name a few).

My specific proposal is based on the traditional experimental items of the "Wason selection task". The task consists in participants being asked to turn those cards, one face of which represents an antecedent A and the other face representing a consequent B, that could disconfirm a given conditional "If A, B". I will assume as given that the modal responses of participants (see Samuels et al (1999)) can be readily explained based on a combination of the explanatory devices provided in chapters 2 to 4.

Specifically, a symbolic version of the task is responded to according to the interpretation of the conditional as per chapter 2 whilst responses to deontic versions of it are aligned with social acceptance

[23] This being said, the thought experiment provided in section 5.2 below is more of a metaphor as a thought experiment. The interested reader is referred to Jo Bervoets (2005d) "To the possible via the desirable" for a fully worked out thought experiment, which did not fit here for reasons of brevity.

(chapter 3) of basic deontic conventions (chapter 4), in the form of a "cheater detection" element of common sense. In the former case, against Johnson-Laird (1993), there is no need for mentally working out mental models. In the latter case, with Sperber et al (2002), the responses need not imply any kind of specific neuronal circuitry leading to a "cheater detection module" on which human reasoning would be based (as parasitic on its proper function of detecting cheaters).

The independent variable of the proposed experiment then is also not of the traditional form that gave rise to most of the traditional experimental items in this field. Changes in response correlated to change in framing, content or context of the task are not indicative, on my theory, of anything psychological, let alone anything neuronal. Rather, they indicate differences in cultural practice in different contexts and with different contents. If anything they are a matter for anthropology (or philosophy, if dependent on objective a priori grounds, if a matter of progress in pure reason). In some cases (for instance the frequentist variations on probabilistic tasks as per Gigerenzer (1999) they simply are different tasks (as per section 3.2.1).

The proposed independent variable instead consists in varying the circumstances of the task, rather than to the material of the task. The prediction, following my over-all conjecture, is that in the suitable task circumstances, participant responses will converge to the "conventional" normative view of logic. This approach is not wholly without precedence in the literature: as per Ford & Billington (2000):

"There are other questions about human nonmonotonic reasoning that our study suggests. For example: Do people fare better if you give them graphical representations like those we have given in this paper? Do people fare better if you give them premises represented in Euler circles, like those used by some of our subjects?"

Obviously, also since within the work of Ford there is quite some stress on "within subjects" type of explanations, the specific example of Euler circles (or similar) has all of the potential "leading the witness" issues as highlighted by Kahneman (2003). We need a rather more gentle variation than, for instance, giving the truth tables of the classical interpretation of the conditional in propositional logic. It is, after all, not our aim to discover individual differences in an ability to work with advanced tools (as is probably the correct interpretation of Stanovich & West (2000)) but to discover a common tendency of humans to "conventionalize" their thought along classical "normative" and truth-functional lines in a marked dissociation from their individual intuitive responses as empirically found in these tasks.

It is then my specific prediction that when participants are induced to reason "conventionally", they will tend to answer in line with classical normative theories. Indeed, these theories are the first to have been discovered and therefore the easiest internalized via common sense (far easier in any case than the advanced explanations considered in chapters 2 and 3). Obviously, the challenge of such an experiment lies in defining a practicable variation of task circumstances in line with the present conjecture but with no implication of a specific solution strategy. I believe one should see successive convergence (which is, as it should be with such a basic convention as the truth-functional conditional, largely independent of cultural background) to a more conventional or "normative" answer along the variation scale below:

- Dropping remarks to the effect that there are no "good or bad" answers,
- Announcing that participants need to justify their answer to a panel afterwards (whilst maybe of some informal use (see Ford (2004) and Donovan & Epstein (1997), the actual justification of responses given afterwards is of no formal value – we are not testing explicit knowledge of justification, not even the ability to come up with it,
- Providing paper and pencil to the participants (with or without motivating then to use it – I'd predict the actual use of it would heavily depend on elements such as above) such that they're able to "simulate" via trial and error there results,
- Conducting the experiment with small groups (without the need to give "one" group answer, it is not a test of negotiation techniques), although this obviously introduces many hidden types of variables, or,

- Providing the task at different early developmental stages. I predict "fallacious" response rates correlate negatively with age following social exposure to advanced conventional argument.

Obviously, the smaller the variation needed to induce the predicted result, the higher the explanatory power of the experiment. The experiment, for many reasons but at least to be comparable to a baseline mentioned above, should be strictly "between subjects".

5.2. The story of a carpenter:

Once upon a time there was a carpenter. She learned to craft tools that allowed her to make something out of wood that was quite useful. She called it a door. After some time she discovered that measuring the hole where the door was to come she could, within bounds of lengths and materials used, produce a matching door. She understood the grammar of doors and quickly found that it applied to windows and a whole suit of other interesting things. But while doors and windows were pretty straightforward items to make, she soon found that applying her knowledge to houses was less straightforward. She needed a helping hand from her friend in writing down the rules of her grammar as well as a practical constraint here and there that soon became evident after a few failures that succumbed under their own ambition – after which she was able to construct a becoming home to the leader of her group.

At this point in time the wise man, the lifelong aide to the leader, became suspicious. Some say it was a case of jealousy but little is known about his motivation as the wise man spoke only seldom and if he spoke he only spoke privately to the leader. The wise man accused the carpenter of using witchcraft, in building such houses relying on unnatural magical tools and numbers. The king was taken aback but he settled on a challenge the wise man had devised for the carpenter. She was to build a house without the help of her magical tools and her magical numbers. She was to show that the ability to construct was of her own and did not derive from unreal help from the outside.

She failed and was jailed in her magical house.

The discussion on (ir)rationality or bounded rationality from psychology is like the advise of the wise man. It stems from the fruitful fallacy equating rationality with mind and mind with brain. It is, to me, reminiscent of the Scholastic days where everybody disagreed based on an agreement that one was able to rely on pure reason alone to defend one's metaphysical position.

Whilst explaining the above proposed experiment (assuming, maybe somewhat optimistically, that it'll confirm my prediction) would be feasible through positing 2 types of reasoning (intuitive and analytic as per a.o. Donovan & Epstein (1997), these types of thought experiment could, in my view, show that such an interpretation is not really a tenable one. A more advanced thought experiment would of course include in a more direct way than in the above the fact that accepting such a "2-system" view of human reasoning leads to one or more of the antinomies informally discussed in chapter 1.

What remains to be done, and will have to be done in a very sketchy way, then is explaining how come the intuition with respect to this "2-system view" – the phenomenology of conscious and subconscious – is so widespread. This is done in the next, and final, chapter.

Chapter 6 – The origin of consciousness as a mirror of society:

" It is like asking a flashlight in a dark room to search around for something that does not have any light shining upon it. The flashlight, since there is light in whatever direction it turns, would have to conclude that there is light everywhere.
And so consciousness can seem to pervade all mentality when actually it does not."

J. Jaynes, "On the origin of consciousness in the breakdown of the bicameral mind", 1976.

Having answered the title question in the negative, at least insofar machines are perceived as thinking (functionally making computations over internal representations), I am at odds with the long-standing tradition in the humanities seeing humans inter alias as "thinking things". So much so that I seem to deny the very truth Descartes took that as the ultimate, and only, certainty. The question seems more than fair – although not quite crucial as to the merit of the rest of this thesis – how I can defend an account as the present in the face of the phenomenologically undeniable "stream of consciousness" and how I can suppose it tells us more about the human psyche than a prima facie fact of having some kind of Turing-machine in my head, sequentially crunching data to come up with analytic solutions.

I hope I have shown, or at least have made enough of a start with it to allow other people to show, that a line of reasoning as per the end of chapter 4 should allow us to build a defence. With the procedure as indicated in chapter 4 together with the devices explained in chapters 2 and 3, it should not be too far-fetched to construct very complex behaviour from relatively unsophisticated initial, native, material. In fact, complex behaviour and mental phenomena need not be determined at all by native material, as per Gibson (2001), the present conjecture allows the materials to be as complex as the world is with as only condition allowing the priority question to be "what is outside of the head?" instead of asking "what is inside it?".

From the bottom-up, we need not posit anything sublime that is given except some basic circuitry (as a matter of fact, lots of it) to implement perception, conditionals as well as selective attention in which to carry out arbitration as to what is more relevant given the current situation. The only catch is that, in so doing, we need not kid ourselves in maintaining the ambition that our brains somehow would be able to allow reconstruction of complex pure argument. As depressing as this may seem to some, no great loss is incurred by giving up this ambition. Complex, pure argument can be constructed socially. Not only can it be maintained in books but it can even be embodied in beings entirely different than ourselves in one crucial respect: machines that think.

All of this was inspired by J. Jaynes (1976) where I first read the basic conjecture that consciousness is not a function of some part of the brain – whether pineal gland or module – but, conceivably imitation of actual external social behaviour via a stream of consciousness parasitic on working memory (when it handles conditionals related to sufficiently complex social behaviour, see chapter 4). I hope I have done more to serve this basic idea by working bottom-up instead of from the heights of ancient texts. It is of course inherently confusing when simple devices give rise, together with experience, to behaviour as in consciousness. One easily mistakes the ability to come up with answers to complex problems with the ability to implement complex solutions for problems. As the quote above tries to say, one also easily confuses the feeling of reasoning steps in one's head for the knowledge that there is a reasoning device in one's head. Such is, I believe, the basic confusion underlying mental model (Johnson-Laird (1993)) and similar, but syntactic, theories. This feeling is not more than a mirror of how people reason in the social, critical circumstances of today (and is limited to the classical 4-6 items of working memory).

So, we are not "blockheads" with a pyramid of reason inside our heads (some pyramids being higher in the intelligent at the expense of emotionality and perception, or vice versa). We are all to some extent part of a collectively constructed pyramid of reason which we can observe but never internalize. So much so that the Habermasian discourse ethics (Habermas (1991)) is quite literally grounded upon common sense: our individuality only comes as a result of our social context. Each individual is maybe the most precious thing that can be but all of that preciousness is not his only as it derives from his interplay with the others.

In order not to close on a note almost entirely foreign to the substance of this thesis, let me quote Julian Jaynes one more time. This time it is a quote that should avoid people to impute anything mystical to theories as the present one (see Jaynes (1976)):

"Let us not make a mistake. When I am conscious, I am always and definitely using certain parts of my brain inside my head. But so am I when riding a bicycle, and the bicycle riding does not go on inside my head. The cases are different of course, since bicycle riding has a definite geographical location, while consciousness does not. In reality, consciousness has no location whatever except as we imagine it has."

I thank you for your persistence in reading this.

IV. Bibliography

To Gene Roddenbury, here present in spirit as one of the greatest philosophers that ever lived.
He boldly went where no man ever went before.
He went to a place where it didn't matter of what you were made.
Where the only thing that does matter is the ability to communicate with each other.
That place, I hope, is this place.

E. W. Adams (1975), "The logic of conditionals", D. Reidel, Dordrecht.

E. W. Adams (2002), "On a proportionality analysis of syllogistic reasoning", Proceedings of the 1st Salzburg Workshop on Paradigms of Cognition, Salzburg.

A. D. Baddeley (1999), "Essentials of human memory", Psychology press, Taylor & Francis Group.

L. W. Barsalou (1999), "Perceptual symbol systems", Behavioral and Brain Sciences 22, pp. 577-660".

S. Benferhat, J. F. Bonnefon & R. M. Da Silva Neves (2002), "An overview of possibilistic handling of default reasoning: applications and empirical studies", Proceedings of the 1st Salzburg Workshop on Paradigms of Cognition, Salzburg.

J. Bervoets (2005a), "Symbolic perceptual systems?", paper submitted for the course "Computational and modular approaches in cognitive science" given by J. Wagemans, Katholieke Universiteit Leuven.

J. Bervoets (2005b), "Information as feedback", paper submitted for the course "Computational and modular approaches in cognitive science" given by J. Wagemans, Katholieke Universiteit Leuven.

J. Bervoets (2005c), "Three rules for the game of categorization and induction", paper submitted for the course "Cognitive Modeling of categorization and induction" given by G. Storms, Katholieke Universiteit Leuven.

J. Bervoets (2005d), "To the possible via the desirable", paper submitted for course "Computational and modular approaches in cognitive science" given by J. Wagemans, Katholieke Universiteit Leuven.

I. Bratko (2001), "PROLOG. Programming for Artificial Intelligence", 3d Edition, Pearson Education Limited.

P. W. Cheng & K. J. Holyoak (1985), "Pragmatic reasoning schema's", Cognitive Psychology, Vol. 17, pp. 391-416.

G. Chierchia & S. McConnell-Ginet (1996), "Meaning and Grammar. An introduction to semantics", MIT press.

D. Dennett (1994), "The role of language in intelligence", in "What is intelligence" (edited by Jean Khalfa), Cambridge University Press.

K. Dieussaert (2002) – KD, "On defeasible inferences", dissertation to obtain title of Doctor in Psychology, Katholieke Universiteit Leuven.

Donovan & Epstein (1997), "The Difficulty of the Linda Problem can be attributed to its simultaneous concrete and unnatural presentation, and not to conversational implicature", Journal of Experimental Social Psychology 33, 1-20.

D. Edgington (1995) – DE1, "On Conditionals", Mind, New Series, Vol. 104, No. 414, pp. 235-329.

D. Edgington (2001) – DE2, "Conditionals", in "The Blackwell guide to Philosophical Logic", Edited by Lou Gobles, pp. 385-414, Blackwell.

R. Elio & F. J. Pelletier (1997), "Belief change as propositional update", Cognitive Science, Vol. 21, No. 4, pp. 419-460.

J. St. Evans (1991), "Theories of human reasoning: the fragmented state of the art", Theory and Psychology, Vol.1, No 1, pp. 83-105.

J. St. Evans (1993), "Bias and rationality", Chapter 1 in "Rationality: Psychological & Philosophical Perspectives", Edited by K. I. Manktelow and D. E. Over, Routledge.

Fodor (2000), "The mind doesn't work that way" (the scope and limits of computational psychology), MIT press.

M. Ford & D. Billington (2000), "Strategies in human nonmonotonic reasoning", Computational Intelligence, Vol. 16, pp. 446-468.

M. Ford (2004), "System LS: a three-tiered nonmontonic reasoning system", Computational Intelligence, Vol. 20, No, 1, pp. 89-108.

J. J. Gibson (2001), "The theory of information pick-up and its consequences", Reading 4 in "Visual Perception. Essential Readings" edited by S. Yantis, Sheridan Books.

G. Gigerenzer & R. Selten (1999) (Editors), "Bounded rationality. The adaptive toolbox", Report of the 84th Dahlem workshop, MIT press.

L. Gobles (editor) (2001), "The Blackwell Guide to philosophical logic", Blackwell Publishing.

H. P. Grice (1989) – HPG, "Studies in the way of words", Harvard University Press.

J. Habermas (1991), "Erläuterungen zur Diskursethik", suhrkamp taschenbuch.

J. Jaynes (1976), "On the origin of consciousness in the breakdown of the bicameral mind", Mariner books.

P. N. Johnson-Laird & R. M. J. Byrne (1993), "Models and deductive rationality", Chapter 8 in "Rationality: Psychological & Philosophical Perspectives", Edited by K. I. Manktelow and D. E. Over, Routledge.

P. N. Johnson-Laird (2001), "Mental models and deduction", TRENDS in cognitive science, Vol. 5, No. 10, pp. 434-442.

D. Kahneman (2003), "A Perspective on Judgment and Choice" (Mapping Bounded Rationality), American Psychologist, Volume 58, pp. 697-720.

I. Kant (1996), "Kritik der reinen Vernunft", Philip Reclam, Stuttgart.

Lloyd K. Komatsu (1992), "Recent views of conceptual structure", Psychological Bulletin, Vol. 112-3, pp. 500-526.

Henry E. Kyburg Jr., Choh Man Teng (2001) – HK, "Uncertain Inference", Chapters 9-12, Cambridge University Press.

E. J. Lowe (1993), "Rationality, deduction and mental models", Chapter 8 in "Rationality: Psychological & Philosophical Perspectives", Edited by K. I. Manktelow and D. E. Over, Routledge.

D. Marr (2001), "The philosophy and the approach", Reading 5 in "Visual Perception. Essential Readings" edited by S. Yantis, Sheridan Books.

D. L. Medin, J. D. Cole, Gert Storms, B. K. Hayes (2003) – DM, "A relevance theory of induction", Psychonomic Bulletin and Review, Vol. 10, pp. 517-532.

Oaksford & Chater (1993), "Reasoning theories and bounded rationality", Chapter 2 in "Rationality: Psychological & Philosophical Perspectives", Edited by K. I. Manktelow and D. E. Over, Routledge.

Daniel N. Osherson, Edward E. Smith, Ormond Wilkie, Alejandro López, Eldar Shafir (1990) – DO, "Category-based induction", Psychological Review, Vol. 97-2, pp. 185-200.

Papathomas (1999), "The brain as a hypothesis- constructing-and testing agent", Chapter 7 in "What is cognitive science", Blackwell Publishers.

K. Popper (1979), "Objective Knowledge. An evolutionary approach", Revised Edition, Clarendon Press.

H. Rott (2001), "Change, choice & inference. A study of belief revision and nonmonotonic reasoning", Clarendon Press.

I. A. Sag, T. Bender, E. M. Bender (2003), "Syntactic theory. A formal introduction", 2nd edition, CSLI Publications.

Samuels et al. (1999), "Rethinking rationality: from bleak implications to Darwinian modules", Chapter 3 in "What is cognitive science", Blackwell Publishers.

Steven A. Sloman (1993), "Feature-based induction", Cognitive Psychology, Vol. 25, pp. 231-280.

D. Sperber (1996), "Explaining culture. A naturalistic approach", Blackwell Publishing.

D. Sperber & D. Wilson (1986), 'Relevance: Communication and Cognition", Blackwell publishing.

D. Sperber et al. (2002), "Use or misuse of the selection task?" (Rejoinder to Fiddick, Cosmides, and Tooby) *Cognition, Volume 85, Issue 3, Pages 277-290*

Stanovich & West (2000), "Individual differences in reasoning" (Implications for the rationality debate), Behavioral and Brain Sciences, 23, pp. 645-665

D. Wilson & D. Sperber (1979), "Ordered entailments: an alternative to presuppositional theories", in "Syntax and semantic"edited by W. Oh and D. Dineen., pp. 229324.

L. Wittgenstein (1998), "Culture and value", Blackwell Publishing.

Yama (2001), "Matching versus optimal data selection in the Wason selection task", Thinking and Reasoning, 7 (3), pp. 295-311.