# The New Logic

DOV GABBAY, *Department of Computer Science, King's College, Strand, London, ENGLAND, WC2R 2L2, e-mail: dg@dcs.kcl.ac.uk*

JOHN WOODS, *The Abductive Systems Group, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, CANADA, T1K 3M4, e-mail: woods@uleth.ca*

## Abstract

The purpose of this paper is to communicate some developments in what we call the new logic. In a nutshell the new logic is a model of the behaviour of a logical agent. By these lights, logical theory has two principal tasks. The first is an account of what a logical agent is. The second is a description of how this behaviour is to be modelled. Before getting on with these tasks we offer a disclaimer and a warning. The disclaimer is that although the new logic is significantly different from it, we have no inclination to see the new logic as a *rival* of mathematical logic. We do not advocate the displacement of, e.g. model theory, but rather its supplementation or adaptation. The warning is that, whereas mathematical logic must eschew psychologism, the new logic cannot do without it. The fuller story of the new logic is part of our book, *The Reach of Abduction*, scheduled to appear in 2001 or early 2002.

## 1 The Sort of Thing a Logical Agent Is

### *1.1 Psychologism and the New Logic*

Is there a logic of discovery? Some say not. Others are not so sure. We ourselves are in a third camp: In work underway we are actually trying to codify such a logic. Critics of the logic of discovery, those who think it a misbegotten enterprise as such, are frequently drawn to the idea that accounts of how people entertain and select hypotheses, form and deploy conjectures, and more generally how they think things up, are a matter for psychology. Underlying this view is something like the following argument. Let $K$ be a class of cognitive actions. Then if $K$ possesses an etiology (i.e. is subject to causal forces), this precludes the question of the performing or disperforming the $K$-action for good or bad reasons. If there *were* a logic of $K$-action it would be an enquiry into when $K$-actions are performed rationally, that is, for the right reasons. Hence there can be no logic of $K$.

Against this, Donald Davidson is widely taken as having shown that far from reasons for actions precluding their having causes, reasons *are* causes, or more strictly, *having a reason* for an action is construable as a cause of it. (Davidson [14]. See also Piestroski [54].)

We ourselves are inclined to emphasize a substantial body of work in reliabilist and other forms of causal epistemology. In its most basic sense, a subject performs a cognitive task rationally when his performance of it is induced by causal mechanisms that are functioning reliably, that are functioning as they should. If a logic can be seen as a theory of rational cognitive performance, the present argument fails to demonstrate the impossibility of a logic of discovery. Even so, the idea of logic as a

 

theory of rational performance runs into a different, though related, objection. The trouble with such a view of logic, it is said, is that it commits us to *psychologism*, and psychologism is false.

Anti-psychologism is not a single, stable thesis. It is at least three theses, pairwise inequivalent.

1. In one sense, it is the case made by the argument we have just re-examined and rejected.
2. In another sense, it is the view that although logic deals with the canons of right reasoning, no law of logic is contradicted by any psychological law or psychological fact.
3. In a third and more emphatic sense, it is the view that logic has nothing whatever to do with how people do reason or should.[1]

Having dealt with anti-psychologism in the first sense, it remains to say something about the other two. Sense number two need not detain us long. It is a view of anti-psychologism which is accepted by logicians who take a traditionally normative view of logic. On this view, psychology is purely descriptive, and logic is purely prescriptive. Hence the laws of logic remain true even in the face of massive misperformance on the ground. On the other hand, those who plump for reliabalist theories of rational performance will reject anti-psychologism in its present sense, just as they reject it in sense number one.

This leaves the third conception, the idea that logic has nothing to do, normatively or descriptively, with how human beings — or other kinds of cognitive agents, if any — think and reason. It is a view with an oddly old-fashioned ring to it, suggesting a position which seems to have been over-taken by developments of the past quarter century, referred to collectively by the founding editor of the *Journal of Logic and Computation* as the new logic. He writes:

> Let me conclude by explaining our perception of the meaning of the word 'Logic' in the title of this Journal. We do not mean 'Logic' as it is now. We mean 'Logic', as it will be, as a result of the interaction with computing. It covers the new stage of the evolution in logic. It is the new logic we are thinking of. (Gabbay [26].)

Ten years later the present authors noted (Gabbay and Woods [24].)[2] that the editor's prediction has been met with a notable degree of confirmation. Non-monotonic logics, default logics, labeled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics are now better understood and methodologically more secure than they were a decade ago. Imaginative re-examinations of fragments of classical logic have produced fresh insights, including, at times, decision procedures for and equivalency with non-classical systems.

---

[1] It is interesting that the case which Frege actually pressed against psychological methods in logic are not transparently present in the trio of interpretations currently in view. In Frege [21] and subsequent works, Frege's resistance was twofold. First, if psychological methods were engaged in such a way as to make mathematics an experimental science, then those methods should be eschewed. Second, if psychological methods were engaged in such a way that mathematics lost its intersubjective character, then psychological methods should be either abandoned or not employed in such ways. It bears on the present point that whereas Boole was a psychologcist about logic, and whereas Frege was a critic of Boole, Frege never criticzed Boole for his psychologism. Logic for Boole is not a matter of how people *actually* think but rather is a normative account of the correct use of reasoning (Boole [5, pp. 4 and 32]).

[2] This and the following eight paragraphs are adapted from Gabbay and Woods [24] by permission.

One of the more interesting features of the new logic has been research partnerships with fallacy theory, the logic of natural language reasoning and argumentation theory.[3] For many decades the once-active relationship between formal and informal logic had suffered a kind of collapse, which was marked by a general indifference. How this disaffection came about can be explained. The single most influential turn in the history of logic in the past century and a quarter was the appropriation of logic by mathematics. It was an ironic transformation, in as much as the motivating factor in Frege's reshaping of logic was logicism, the thesis which proposes the appropriation of mathematics by logic. Whatever the influences that were in play in 1879, mathematical logic evolved into an impressive maturity in its four domains of proof theory, recursion theory, model theory and set theory. In the interplay of rigour, precision and sheer creativity, the four sectors of mathematical logic set extremely high standards. Borne by these successes, each of the branches of post-Fregean logic was able to set its own research programmes. It is hardly a surprise that Frege's own rationale for mathematical logic should long ago have slid from view, to say nothing of the rationale articulated by Aristotle, the first logician. For Frege, logic had to be re-designed for its role in a proof that arithmetic was expressible without relevant loss in pure logic (which for Frege also included a kind of set theory). Aristotle's interest in logic was entirely different. Logic was required to produce a theory of syllogisms, and syllogistic logic would be the indispensable theoretical core of a general theory of argument and strict reasoning.

As it has evolved, mathematical logic displays no serious loyalty to logicism (if anything, the opposite is true), and apart from lazily misguided assurances to First Year Logic students that first order quantification theory will reveal the secrets of good argument and inference, mathematical logic has none of the promise of Aristotle's founding conception. This is not to say that there isn't the occasional contingent exception to these alienations. For example, the semantics for intuitionistic modal logic helps in modelling the fallacy of *petitio principii*, (Woods and Walton [66, Ch. 10]) the analysis of which in turn validates certain inferences in paleontology heretofore judged fallacious. (Harper [31])

The general separation of informal from formal (mathematical) logic derives mainly from the concurrence of two significant facts. One is that informal logic, including fallacy theory, became, in Lakatos' sense, a degenerate research programme. (Hamblin [30, Ch. 1]) The other is that as this degeneracy worsened, mathematical logic went from triumph to triumph. Even the most mathematically discouraging result that logic is capable of delivering was received as a triumph, and the name of Gödel was enshrined in our intellectual history for all time.

The new logic, the logic born of the application of the technical sophistication of mathematical logic to the project of informal logic, has triggered the very rapprochement that mathematical logic was not structured to deliver or to seek. The new logic, whatever its various differences of mission and detail, has sought for mathematically describable models of what human agents actually do in real-life situations when they cogitated, reflected, calculated and decided. Here was an approach that would in an essential way take what mathematical logic would see as inert context into the theory itself, where it would be directly engaged by the ensuing formalisms.

[3] Attested to by the Netherlands Royal Academy Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

A further attraction of the new logic has been a re-awakening of the research programme in the History of Logic. For most of its long life, logic has pursued missions that would be as recognisable to Aristotle as to many a reader of the *Journal of Logic and Computation*, as well as this journal. Logic had something directly to do with real-life human competence in reasoning and in arguing. This is a mission which mathematical logic can neither disapprove of nor engage, and is not something that any such logician need regret or even care about. Still, if mathematical logic were seen in a hegemonic way, if it were granted sole licence to the name of logic, then (only with some notable exceptions) the History of Logic could only be seen as the history of failure, the quaintness of its antiquarian interest notwithstanding. Not having to endure such constraints, the new logic is free to see in the theoretical labours of our forbears examples of genuine theoretical value, and insights or doctrines of material interest to researchers of the present day. Whether it is the discovery that Aristotle was the first relevant intuitionistic, nonmonotonic, hyperconsistent logician, (Woods [69]) or whether it is the realization that Peter of Spain anticipates results of importance to Cognitive Modelling, the overall effect of this reclamation has been the liberation of logic's history from the status of the museum-piece.

If psychologism is the view that logic has something to do with how beings like us think and reason, then we are psychologicists. But we are psychologicists of an ecumenical bent which favours the theoretical rapprochement of logic more narrowly conceived with cognitive science and computer science. It is an approach to logic which leaves it an open research programme as to whether there might be, for example, a satisfactory logic of discovery.

In so saying, we do not place ourselves squarely in or squarely out of the ambit of our interpretations of psychologism (save the first, where the verdict is *out*). In particular, we have not expressly declared ourselves on what might be called Boole's question. Is our approach one in which how people do reason is ignored in favour of how they should reason? Our answer at this stage is somewhat equivocal, but it is the best we can do for now: we have doubts about the purported exclusiveness of this very distinction. (Woods [71, ch 8])

## *1.2 A Hierarchy of Agencies*

There is an intuitive distinction between the cognitive practices of individuals caught up in the affairs of everyday life, including matters of Vital Importance, as Peirce calls them, and rational strategies in science. Right from the outset the distinction is vexed by a process-product ambiguity attaching to the word 'science'. If we mean by science its formal theory, i.e. the highly structured set of sentences which report its finished results in the theorems it proves, then the contrast has sharp edges. But if we mean by science the joint and several activities of real-life practitioners, there is little to our purported contrast apart from a difference in goals, and — in certain cases at least — subject matter.

In our approach, a logic is a formal and somewhat idealized description of a logical agent. An account of the logic must, therefore, begin with a description of the sort of thing a logical agent is. Much of what we shall touch on here will be manifest in the literature on software agent technology.

We think of agency as a hierarchy of goal-directed, resource-bound entities of vari-

ous types. At the bottom of this hierarchy are individual human beings with minimal efficient access to institutionalized data bases. Next up are individual human beings who operate in institutional environments — in universities or government departments for example — which themselves are kinds of agents. Then too, there are teams of such people. Teams can have both short and long histories. Those with very long histories stretch the common sense meaning of 'team', and we speak instead of disciplines, and of other corporate entities such as the American Space Program or Soviet Science in the 1970s. The hierarchy proceeds thus from the concrete to the comparatively abstract, with abstract structures being aggregations of entities lower down. Interesting as this metaphysical fact might be, it is not the dominant organizing principle of the hierarchy. The organizing principle is economic. Entities further up the hierarchy command resources, more than those below are capable of, and also, often enough, of a kind unavailable to the others.

So conceived, the hierarchy is a *poset*, a cluster, **A**, of objects partially ordered by the relation , **C**, of *commanding greater resources than*. We do not here attempt a proof of the partial ordering. It suits our purposes that it is evident that such an ordering exists and that it has the broadly economic character we are about to attribute to it. It suffices to emphasize that the more highly an agent or agency is placed in the hierarchy the more economic advantages he (or it) commands.

Every agency in this hierarchy $\langle \mathbf{C}, \mathbf{A} \rangle$ involves, whether by aggregation or supervenience or in some other way, the individual agent, the lowest of the low. Such agents are thus basic to the account that we shall give the new logic, and it is to them that we shall concentrate our attention in the present paper.

We see a certain value in identifying *practical* agency with individual agency, and *theoretical* agency with corporate agency, i.e., with agency-types higher up in the hierarchy, the higher the more theoretical. These are not the only defensible equations, needless to say. But they have the virtue of emphasizing a conception of practicality which is fundamentally a matter of the deployment of scarce-resources in what can be thought of as a cognitive economy. (See further Woods et al. [67].)

Like all agents in the hierarchy, the individual is a performer of actions in real time. Some of his actions are physical, such as mowing the lawn. Others, like thinking up a streamlined proof of the completeness of first-order logic, are different. We are not interested at this juncture in joining metaphysical wrangles about the meaning and existence of the mental. 'Mental' is for us a term of commonsense on which we have no specifically theoretical designs.

Nearly everything an individual is faced with doing, or is trying to do, can be done at the wrong time. It can be done at a time so wrong as to court equivalency with not doing it at all, or doing some opposite thing. The person who prepares his next meal makes a huge blunder, assuming no nutritional alternative, if he presents himself with the completed lunch thirty days after initiating the project. It is not enough that agents do the right thing, i.e. perform tokens of the right action-types. It is very often essential that the right thing be done at the right time. However, as we look upwards at this agency-hierarchy, we see a diminished susceptibility to this exigent timeliness. No one doubts that NASA had a real deadline to meet in the sixties, culminating in the moon shot. It might have been that the moon program would have been canceled had that deadline not been met. Even so, individuals are exposed to myriad serious dangers, many of them mortal, that nothing up above will ever know on this scale;

and essential to averting such dangers is doing what is required on time.

The dominant requirement of timeliness bears directly on a further constraint on individual agency. Individuals wholly fail the economist's conceit of perfect information. Agents such as these must deal with the nuisance not only of less than complete information, but with databases that are by turns inconsistent, uncertain, and loosely defined. To these are added the difficulties of real-time computation, limited storage capacity and less than optimal mechanisms for information-retrieval, as well as problems posed by bias and other kinds of psychological effects.

The three great scarcities that the individual must cope with are information, time, and computational capacity. It is precisely these that institutional agents command more of, and very often vastly more of. With few (largely artificial) exceptions, the individual agent is a satisficer rather than an optimizer.[4] For the most part, even seeking to be an optimizer would be tactically maladroit, if not suicidal. The fact of the robust, continuing presence of human agents on this Earth amply attests to their effective and efficient command of scarce resources. It is a fact in which is evident the human capacity to compensate for scarcities of information, time and fire-power.[5]

### *1.3 Scarce-resource Compensation Strategies*

It is evident that the individual agent has the means to flourish in conditions of scarcity, concerning which, we postulate his access to scarce-resource compensation strategies of the requisite kind. Here in rough outline, and no particular order, are the compensation-factors that strike the present authors as particularly important.

Human beings are natural **hasty generalizers**. It was a wise J.S. Mill who observed (Mill [47]) that the routines of induction are not within the grasp of individuals, but rather are better-suited to the resource capacities of institutions. The received wisdom has it that hasty generalization is a fallacy, a sampling error of one sort or another. The received wisdom may be right, but if it is, individual human agency is fallacy-ridden in degrees that would startle even the traditional fallacy-theorist.[6] Bearing on this question in ways that suggest an answer different from the traditional one is the fact that the individual's hasty generalizations seem not to have served his or her cognitive and practical agendas all that badly. Upon reflection, in the actual cases in which a disposition towards hasty generalization plays itself out, the generalizations are approximately accurate, rather than fallacious errors, and the decisions taken on this basis are approximately sound, rather than exercises in ineptitude. Not only is the individual agent a hasty generalizer, he is a hasty generalizer who tends to get things right, or not wrong enough to matter for survival and prosperity.

How is it possible that there be a range of cases in which projections from samples are so nearly right while at the same time qualifying as travesties of what the logic of induction requires? The empirical record amply attests to a human being's capacity

---

[4] An optimizer is an agent who aims for best possible result. A satisficer is an agent who aims at a lesser standard, provided that in the circumstances it is 'good enough'.

[5] It is interesting to contrast the present situation with a simplified logic programming model which operates abductively at zero-cost. In this set up, an abduction problem presents itself. Stripped to essentials, the problem precasts a command in the form, 'Find an $X$, such that $Y$ is derivable'. The no-cost solution is simply to put $X$ as $Y$. It isn't this way in real-life.

[6] On what we are calling the traditional account of fallacies, hasty generalization is always an error. For a contrary view see Woods [68, ch 9].

for *sub-inductive* generalization and projection. It would appear that exercise of this capacity involves at least these following factors, some of them structural, some of them contextual. The sub-inductive generalizer does not generalize to universally quantified conditional propositions. Rather he generalizes to **generic** propositions. There is a world of difference between 'For all $x$, if $x$ is a tiger then $x$ is four-legged' and 'Tigers are four-legged.' The former is falsified by the truth of any negative instance, whereas the latter holds true even in the light of numerous negative instances of certain kinds. We could characterize this difference by saying that universally quantified conditional statements are *fragile*, whereas generic statements are *robust*.[7]

The robustness of what the sub-inductive generalizer generalizes to serves his interests in other ways, two of which are particularly important. One is that the individual agent is *fallible* in (virtually) everything he thinks and does. The other is that the individual agent has the superficially opposite trait of very *high levels of accuracy* in what he thinks and does when operating at the level ordained for him by our hierarchy of agency. Generalizing to generic statements is a way of having your cake and eating it too. It is a way of being right even in the face of true exceptions. It is a way of being right and mistaken concurrently. Generalizing in this way also works a substantial economy into the individual's cognitive effort. It comes from the smallness of its samples and the robustness of its generalizations. Generic inference is inference from small samples under conditions that would make it a fatally stricken induction. We see in this the idea of the affordable mistake. Generic inference is not truth-preserving. One can be wrong about whether Pussy the tiger is four-legged even though one is right in holding that tigers are four-legged. Affordable mistakes are like small infections that help train up the immune system. Just as an infant's summer sniffles is an affordable (in fact, necessary) infection, so too the small errors of the cognitive agent provide him evolving guidance as to the freedom and looseness with which to indulge his predilection for comparatively effortless generalizations. Baby's summer cold loops back benignly in the discouragement of more serious illness. Affordable mistakes likewise loop back benignly in the discouragement of serious error. We can now see that the old saw of *learning from our mistakes* has a realistic motivation. We do not learn from mistakes that kill us.

What is it about such samples that sets them up for successful generic inference? It would appear that the record of generic inference is at its best when samples, small as unit sets though they may be, are samples of **natural kinds**. There has been a good deal of philosophical controversy about whether natural kinds actually exist, about whether the intuitive difference between natural kinds and conventional kinds turns on a principled distinction. Even so, we should not disdain the literature from psychology and computer science in which concepts resembling natural kinds seem to be doing useful work, concepts such as those of *frame* (Minsky [48]), *prototype* (Smith and Medin [60]), *exemplar* (Rosch [56], *schema* (Kitcher [40], Brewer and Nakamura

[7] If a generic claim is not guaranteed to fail in the face of certain kinds of true negative instances, it can hardly be surprising that the truth conditions of generic claims are difficult to specify. Some logicians are drawn to a *generalized quantifiers* approach in which 'Tigers are four-legged' is elliptical for 'Most tigers are four-legged.' (See, e.g. Sher [59] on generalized quantifiers.) We ourselves doubt the equivalence. 'Most of John's students are women' could be true without it also being true that 'John's students are women.' (See also Carlson and Pelletier [9] for valuable papers on genericity.) We also note that the failure of a single true negative instance or counterexample to falsify is not solely a matter of generic inference. Theories or databases subject to *approximate* truth-predicates can have this feature too. (Kuipers [41].)

See also Woods and Walton [66] for a not wholly sympathetic treatment of the traditional view generally, and Woods [68] for an even less sympathetic approach.

[6]) and *script* (Schank and Abelson [58]). There is ample evidence to suggest that our classifications are primitive devices of *type-recognition* together with the mechanisms of fight and flight. It is significant that some of our most successful and most primitive inferences involve the recognition of something as dangerous. Generic inference is part and parcel of such strategies. Just as our capacity for recognizing natural kinds exceeds the comparatively narrow range of immediately dangerous kinds, so too does our capacity for generic inference exceed the reach of fight-flight recognition triggers. But whether in fight-flight contexts or beyond, natural kinds and generic inference are a natural pair.

The fallibility of generic inference is also evident in its relation to defaults. A default is something taken as holding, taken to be true, in the absence of indications to the contrary. Most of what passes for common knowledge is stocked with defaults, and generic inferences in turn are inferences to defaults. Default reasoning is inherently conservative and defeasible. Defeasibility is the cognitive price one pays for **conservatism**. And the great appeal of conservatism is also economic. Conservatism is populated with defaults in the form 'X is what people have thought up to now, and still do.' Conservatism is a method of default-collection. It bids us to avoid the cost of fresh thinking, and to make do with what others have thought before us.[8]

Conservatism places a premium on what is already well-received. On the face of it, conservatism is the *ad populum* fallacy run amok. Here too, we might grant the received wisdom (and note the large irony), conceding that individual agents are notorious fallacy-mongers on a scale not dreamed of even by the traditional fallacy theorist. But as we said just above in our examination of a similar indictment of hasty generalization, there are factors which seem to cut across so harsh a condemnation. One is that we are, by and large, enormously well-served by the trust we place in the testimony of others. Popular beliefs are what Aristotle called *endoxa*. They are 'reputable opinions', the opinions of everyone or of the many or of the wise. The mere *fact* of popular opinions triggers an abduction problem. What best explains that $p$ is a proposition believed by everyone? An answer, which certainly can be criticized in respect of certain particular details, but which cannot convincingly be set up for general condemnation, is that $p$'s universal acceptance is best explained by supposing that $p$ is true, or anyhow that a belief in $p$ is *reasonable*. What is loosely called common knowledge is an individual's (or an institution's or a society's) inventory of *endoxa*. What is especially striking about common knowledge is that it is acquired by an individual with little or no demonstrative effort on his own part, and with attendant economies of proportional yield.

It is evident therefore that individual agents depend for what they think and for how they act on the say-so of others, on the more or less uncritical and unreflective testimony of people who by and large are strangers. Here is yet another respect in which the conduct of human agents would seem to fall foul of the received opinion of fallacy theorists (let us not forget that the *endoxa* of the wise, even when they are fallacy theorists, are not *guaranteed* to be true!) For it would appear that individual agents are programmed to commit and implement the program on a large scale, the *ad verecundiam* fallacy. But as before, the natural record of thought and action produced by such dependencies is rather good; most of what we think in such ways is not especially inaccurate and, in any case, not inaccurate enough to have made a

[8] The speed-for-less-than-perfect-accuracy is also evident in the field of quantum computation. See here [15].

mess of the quotidian lives of human individuals. We may suppose, therefore, that the traditional fallacies of hasty generalization, *ad populum* and *ad verecundium* are hardly fallacious *as such* (e.g. when considered as an individual's strategies or components of strategies for practical action), but are fallacies only under certain conditions.

Even so, it is not our view that our disposition to economize in ways *ad verecundiam* is risk-free. In a classic study, Solomon Asch observed that in certain situations of interactive estimation, "whereas the judgements were virtually free of error under control conditions, one-third of the minority estimates were distorted toward the majority." (Asch [1]) On the other hand, as studies of the Delphi Method make clear (Helmer [32]), group opinions on commonplace matters of fact formed by such methods tend to be more reliable than the simple average of the solo judgements of the individual involved.

Human life is dominantly social; individual agents find cooperation to be another thing which, like hasty generalization, is almost as natural as breathing. The routines of cooperation transmit to an individual nearly all of the community's common knowledge that he will ever possess. Even though the complete story has yet to be told, cooperation has received the attention of attractive theories (e.g. Axelrod [2], Coady [11] But see also Gabbay and Woods [23] and Gabbay and Woods [25]).[9]

There is a natural and intuitive contrast between accepting something on the say-so of others and working it out for oneself. Cross-cutting this same distinction is the further contrast between accepting something without direct evidence, or any degree of verification or demonstrative effort on the accepter's part, and accepting something only after having made or considered a case for it. The two distinctions are not equivalent, but they come together overlappingly in ways that produce for individual agents substantial further economies.

Such additional economies are the output of two regularities evident in the social intercourse of agents. One is the **reason rule**:

> *Reason Rule*: One party's expressed beliefs and wants are a prima facie reason for another party to come to have those beliefs and wants and, thereby, for those beliefs and wants to structure the range of appropriate utterances [which] that party can contribute to the conversation. If a speaker expresses belief X, and the hearer neither believes nor disbelieves X, then the speaker's expressed belief in X is reason for the hearer to believe X and to make his or her contributions conform to that belief. (Jacobs and Jackson [38], 57; Jackson [37], 103).

The reason rule reports an empirical regularity in communities of real-life discussants. Where the rule states that a person's acceptance of a proposition is reason for a second party to accept it, it is clear that 'reason' means 'is *taken* as reason' by the second party. Thus a descriptively adequate theory will observe the Jacob's-Jackson regularities as a matter of empirical fact. This leaves the question of whether anything good can be said for these regularities from a normative perspective. If normativity is understood instrumentally, it would appear that the reason rule can claim some degree of normative legitimacy. Not only does it produce substantial economies of time and information, it seems in general not to overwhelm agents with massive error or inducements to do destructive or even silly things. The *reason rule* describes a default. Like all defaults, it is defeasible. Like most defaults, it is a conserver of

[9] But for difficulties with the Axelrod approach, see e.g. Danielson [13].

scarce resources. And like many defaults, it seems to do comparatively little cognitive and decisional harm.

There is a corollary to the reason rule. We call it the *ad ignorantiam* rule:

> *Ad Ignorantiam Rule*: Human agents tend to accept without challenge the utterances and arguments of others except where they know or think they know or suspect that something is amiss.[10]

Here, too, a good part of what motivates the *ad ignorantiam rule* in human affairs is economic. People don't have time to mount challenges every time someone says something or forwards a conclusion without reasons that are transparent to the addressee. Even when reasons are given, social psychologists have discovered that addressees tend not to scrutinize these reasons before accepting the conclusions they are invited to endorse. Addressees tend to do one or other of two different things before weighing up proffered reasons. They tend to accept this other party's conclusions if it is something that strikes them as *plausible*. They also tend to accept the other party's conclusion if it seems to them that this is a conclusion which is within that party's competence to make—that is, if he is seen as being in a position to know what he is talking about, or if he is taken to possess the requisite expertise or authority. (See, e.g. Petty and Cacioppo [52], Eagly and Chaiken [16], Petty, Cacioppo and Goldman [53], Axsom, Yates and Chaiken [3], O'Keefe [50], and the classic paper on the so-called atmosphere effect, Woodworth and Sells [72]. But see also Jacobs, Allen, Jackson and Petrel [39]. We see, once again, the sheer ubiquity of what traditionalists would call — overhastily in our view — the *ad verecundiam* fallacy.)

Part of what a logical agent does is make *abductions*. In that respect, what we might call the *Cut Down Problem* hovers over these pages. How, from so *much* to make use of, does the abductive agent shrink his candidates for hypothetical adoption to so *few*? We see the individual agent as a processor of information on the basis of which, among other things, he thinks and acts. Researchers interested in the behaviour of information-processors tend to suppose that thinking and deliberate action are modes of consciousness. Studies in information theory suggest a different view. It is a view that comports with *Peirce's Principle*, according to which an essential part of an individual's solution of the Cut Down Problem is that he is a successful economizer. (Peirce [51, 5.598–600; 6.528ff; 5.602; 6.529–530; 6.532–538; and 7.221].). It is a view that seizes on a central fact about consciousness.

Inidvidual (or practical) agents come in a standard form. That is to say, in the general case an individual agent is an organic realization of a certain kind of cognitive economy. In this standard form, *consciousness* is a fundamental factor.

Consciousness has a surprisingly narrow bandwidth. It processes information slowly. The rate of processing from the five senses combined—the sensorium, as the Mediaevals used to say—is in the neighbourhood of 11 million bits per second. For any of those seconds, something fewer than 40 bits make their way into consciousness. Consciousness therefore is highly entropic, a thermodynamically costly state for a human system to be in. At any given time there is an extraordinary quantity of information

[10] The inference sanctioned by this rule can be schematized as follows.
1. S is arguing that $p$ is the case.
2. I have no knowledge of anything amiss with this argument.
3. So I accept that $p$.

processed by the human system, which consciousness cannot gain access to. Equally, the bandwidth of language is far narrower than the bandwidth of sensation. A great deal of what we know — most in fact — we aren't able to tell one another. Our sociolinguistic intercourse is a series of exchanges whose bandwidth is 16 bits per second (Zimmermann [73]); it is even less for conversations transacted by telephone.

It is now evident that we must amend the claim that individual agents suffer from a scarcity of information. In so doing however, we are able to lend appropriate emphasis to what remains true about that proposition. In pre- or subconscious states, human systems are awash in information. Consciousness serves as an aggressive suppressor of information, preserving radically small percentages of amounts available pre-consciously. To the extent that some of an individual's thinking and decision-making are subconscious, it is necessary to postulate devices that avoid the distortion, indeed the collapse, occasioned by information-overload. Even at the conscious level, it is apparent that various constraints are at work to inhibit or prevent informational surfeit.

Human agents make do with scarce information and scarce time. They do so in ways that make it apparent that in the general case they are disposed to settle for *comparative* accuracy and comparative *sensibleness* of action. These are not the ways of error-avoidance. They are the ways of fallibilism. Error-avoidance strategies cost time and information. The actual strategies of individual agents cannot afford the costs and, in consequence, are risky. As we now see, the propensity for risk-taking is a structural feature of consciousness itself. It might strike us initially that our fidelity to the reason rule convicts us of gullibility and that our fidelity to the *ad ignorantiam* rule shows us to be lazily irrational. These criticisms are misconceived. The reason rule and the *ad ignorantiam* rule are strategies for minimizing information over-load, as is our disposition to generalize hastily.

Consciousness makes for informational niggardliness. This matters for computer simulations of human reasoning. That is, it matters that there is no way presently or foreseeably available of simulating or mechanizing consciousness. Institutional agencies do not possess consciousness in anything like the sense we have been discussing. This makes it explicable that computer simulations of human thinking fit institutional thinking better than that of an individual. This is not to say that nothing is known of how to proceed with the mechanization of an individual's conscious thinking. We know, for example, that the simulation cannot process information in quantities significantly larger than those recently mentioned.

For individual agents it is a default of central importance that most of what they experience, most of what is offered them for acceptance or action, stands in no need of scrutiny. Information-theoretic investigations take this point a step further in the suggestion that consciousness itself is a response to something disturbing or at least peculiar enough to be an interruption, a demand so to speak to pay attention. If this is right, consciousness is an aberrant state, the exception rather than the rule, and the same is true both for case-making and for the consideration and evaluation of cases. This affects abduction in an especially interesting way, for it squares with the intuition that an abduction problem requires a trigger and that a trigger is an event or state of affairs or scrap of information which stands out in some way, which demands attention and calls for an explanation.

Most of the information processed by an individual agent he will not attend to, and

even if it is the object of his consciousness he will attend to it in as little detail as the exigencies of his situation allow. Arguing is a statistically nonstandard kind of practice for human agents, but even when engaged in it is characterized by incompletions and short-cuts that qualify for the name of enthymeme. The same is true of reasoning, of trying to get to the bottom of things. In the general case, the individual reasoner will deploy the fewest resources that produce a result which satisfies him, which, for example enable him to achieve his goals. Here is further evidence that individuals display a form of rationality sometimes called 'minimal', and well-discussed in Cherniak [1986]. In addition to features already discussed in this chapter, the minimal rationalist is a non-monotonic reasoner, and in ways that are mainly automatic, the successful manger of belief-sets and commitment-sets that are routinely inconsistent. Much of what makes for the inconsistency of belief-sets comes from the inconsistency of deep memory storage and further aspects of inconsistent belief-sets flow from the inefficiencies of memory retrieval.

Conscious is a controversial matter in contemporary cognitive science. It is widely accepted that information carries negative entropy. Against this is the claim that the concept of information is used in ways that confuse the technical and common sense meanings of that word, and that talk of information's negative entropy overlooks the fact that the systems to which thermodynamic principles apply with greatest sure-footedness are closed, and that human agents are hardly *that*.

The complaint against the over-liberal use of the concept of information, in which even physics is an information system (Wolfram [65]), is that it makes it impossible to explain the distinction between energy-to-energy transductions and energy-to-information transformations. Also singled out for criticism is the related view that conscious arises from neural processes. We ourselves are not insensitive to such issues. They are in their various ways manifestations of the classical mind-body problem. We have no solution to the mind-body problem, but there is no disgrace in that. The mind-machine problem resembles the vexations of mind-body, both as to difficulty and to type. We have no solution to the mind-machine difficulty. There is no disgrace in that either.

Indispensable to agency is the ability to remember. The literature on **memory** recognizes a contrast between *occurrent* and *dormant* memory, echoing a distinction between short term and long term memory (for the classic studies see Howe [1970], Collins and Quillum [1969] and Lindsay and Norman [1977].) These two reminiscential operations work in interestingly different ways. Occurrent memory presents beliefs that are here and now, ready for action, for driving inferences and shaping behaviour. On the other hand, beliefs stored in dormant memory are not accessible as premisses in inferences; they do not conflict with or interact with one another; and some researchers are of the view that they do not influence conduct.

Occurrent memory is governed by sharper requirements than dormant memory. Occurrent or short term memory is in some sense bothered by inconsistency, whereas inconsistencies in dormant memory are virtually inert. This difference also crops up in the following way. Occurrent inconsistency is something a rational agent will, in one way or another, try to do something about. Dormant inconsistencies tend not to register in an agent's consciousness. By and large there is nothing to be done about them.

Recent work in the dynamic logics of reasoning (e.g. van Benthem [61]) mark a

distinction between inferences that are dependent on short-term representation and those that hinge on long-term memory, which involves the processing of representations of greater abstraction. So far, the best formal treatments of this kind of modularity in information representation are to be found in theories of *abstract data types*, developed by computer scientists. Inductive logic has had little to say of the developments; but a significant relaxation of logic's traditional deductive boundaries may be found in the more flexibly formatted *labelled deductive systems* (Gabbay [28]) as well as in dynamic logic. (See also Gabbay [27].)

Beliefs and memories are not the only things held to consistency assumptions in the lives of individual agents. Desires are also commonly expected to be consistent: 'beliefs and desires can hardly be reasons for action unless they are consistent.' (Elster [18], 4). In present day science, consistency is often defined for *preferences*. Transitivity is the minimal condition on preference. If agent $S$ prefers $X$ to $Y$ and $Y$ to $Z$ then he can be expected also to prefer $X$ to $Z$.

## 2 The Sort of Thing a Logic Is

### *2.1 Logic as a Description of a Logical Agent*

The structure of minimal rationality shows the individual agent to be the organic realization of a nonmonotonic paraconsistent base logic. There is little to suggest that the strategies endorsed by classical logic and most going nonstandard logics form more than a very small part of the individual agent's repertoire of cognitive and conative (decision-making) skills. 'Putting this more generally, deductive logic so far has little to say about the meso- and macro-levels of reasoning, which is where most of our strategic thinking takes place.' (van Benthem [62], 33.) If it is true, as suggested above, that individuals are in matters of non-demonstrative import sub-inductive rather than inductive agents, the same would also appear to be the case as regards deduction. If so, human individuals are not the wet-wear for deductive logic, at least in the versions that have surfaced in serious ways in the sprawling research programmes of modern logic. There is a particularly interesting reason for this. If we ask what the value of deductive consequence is, the answer is that it is a guarantee of truth-preservation. Guaranteed truth-preservation is a guaranteed way of avoiding error.[11] But individual agents are not in the general case dedicated to error-avoidance. For the most part, the routines of deduction consequence do not serve the individual agent in the ways in which he is disposed (and programmed) to lead his cognitive and decisional life. This is not to say that agents do not perform deductive tasks even when performing on the ground level of our hierarchy of agency-types. There is a huge psychological literature about such behaviour. (Accessibly summarized in Manktelow [46].) The point rather is that deductive thinking is so small a part of the individual's reasoning repertoire.

Let us briefly take our bearings: Complexity is a relatively recent item on the agendas of logicians. It is known that the most extreme complexity embedded in any formal or logical apparatus utterly pales in comparison to the speed with which individual agents perform their cognitive tasks in real time. We have been suggesting a certain explanation of this. The basic idea is that speed is a trade-off for strict soundness and

[11] That is, of avoiding errors not already in the agent's database or premiss-set.

completeness. While cognitive strategies employed by individuals cannot pretend to ensure complete accuracy, still less absolute certainty, they serve us well when things go awry and start to degrade. The kind of cognitive competence which such procedures serve rather well, has nothing to do with the hell-bent accumulation of logical truths or with the output of some well-constructed and well-programmed theorem-prover, but with timely, composed, and sensible reactions to difficulty and challenge. On this view, 'rationality is repair.' (van Benthem [62], 42). The rationality-is-repair approach does not however preclude the possibility of building formal systems with greater real-time fidelity. It is more easily said then done. Van Benthem points out that the logic of refutations and first-order logic have been reformatted for Arrow Logic and Modal First-Order Logic. (Venema [63] and van Benthem [61].) This raises the possibility that decidable systems might in turn be reduced to less complex systems, which might better model real-time cognitive performance. But, a warning: such systems will nevertheless be highly complex.

A third option bears rather directly on a question of how could we write rules for what is largely instinctual behaviour. The present option suggests an answer. It is to construct architectures which represent automatic, subconscious, sublinguistic and (probably) highly connexionest delivery systems for much of what passes for execution of the rationality-is-repair model of cognitive competence. One virtue of this, approach is that the theory has principled occasion to explain why our overt cognitive output, while often wrong in detail, is basically right.

A logic appropriate for the individual agent, a logic of which he can reasonably be said to be a psychophysical realization, will be one that reflects, among other things, his explanatory and interpretive practices in a principled way. In work underway, we show how the combined factors of relevance and plausibility bear on such practices. For the present it suffices to note their crucial involvement with minimal rationality.

Information-theoretic studies of consciousness suggest that the basic structure of consciousness is such as to exclude from his attention most of the information that an individual is processing at any given moment. This in turn suggests a certain approach to the Cut Down Problem. It appears that discounted information is irrelevant to whatever a conscious agent is currently attending to, that consciousness *itself* is a relevance-sieve. Even within consciousness, individuals have the uncanny ability to distinguish the irrelevant from the relevant. Consider an event that has penetrated an agent's consciousness. Already an economically and informationally aberrant occurrence, it stands out in ways that call for attention. In many cases such occurrences call for explanation. For any such occurrence the number of possible explanations is indefinitely large. The number of possible explanations which the individual will actually attend to is correspondingly very small. Thus the **candidate space** of an abduction problem is a small proper subset of an indefinitely large set of possible explainers (or, more generally, possible resolvers). This suggests an operational characterization of relevance. A possible resolver is relevant to an agent's abduction task if and only if it is a member of his candidate space, if and only if it is a possible resolver that he actually considers.

On this account, relevance is indeed a largely automatic affair, which is where the principal economies lie. It is a concomitant of the consideration of possibilities. **Relevance** marks the boundary between possible resolvers and *candidate-resolvers*. It also marks the boundary between the more general distinction between *mere* and

*real* **possibilities**. Something is a mere possibility for an individual agent when it does not intrude itself into the agent's action plan. Mere possibilities are those that give the agent no grounds, proactively or retroactively, for action or for deliberation. Something is a *real* possibility for an agent when and to the extent that he is prepared to give it standing (even counterfactual standing) in his deliberations. An agent might be got to concede that there *might* be a massive earthquake in London later this afternoon. It is a mere possibility for him if the agent gives it no standing in his action-plans for today. It is a real possibility if it is something he is prepared to reflect upon in organizing his day, to reflect upon even if it subsequently meets with his dismissal upon reflection. Like sets of possible explainers, an agent's totality of mere possibilities is a large set at any given time. Like an agent's candidate spaces, his real possibilities constitute a small set at any given time. Just as *relevance* is defined over sets of possible resolvers as that which screens possible explainers into candidate spaces, relevance is likewise the sieve that takes possibilities into real possibilities.

Because the routines of irrelevance-avoidance are mainly automatic, they operate with considerable economy. A further factor in the cognitive and decisional economies of human individuals is the (again largely automatic) command of the distinction between what is plausible and what is not. The distinction is marked by an ambiguity which the would-be theorist should try to keep in mind. At one level, **plausibility** and implausibility are attributes of events or states of affairs. An individual agent may find it implausible that his business partner is a closet neo-Nazi, even if it turns out that this is precisely what he is. His being so could still qualify as implausible. Our distinction also bites at the level of explanation. If the business partner is in fact a neo-Nazi, his colleague may wish to know why. If it is suggested that the partner's neo-Naziism arose from his besottedness with his very beautiful, charismatic and rather imperious girl-friend, this might strike us as a plausible explanation. Or not; and therewith a problem for the theorist. The contrast between the plausible and the implausible links in an important way with an agent's estimate of what *would* and *would not* be the case with regard to the subject of the judgement of plausibility or implausibility. Harry, the neo-Nazi, *would* be the sort of person to succumb radically to the charms of a lover. It is something he would do, where others in the movement came to this neo-Naziism differently. Imbibing it from a girl-friend is not something that *they* would do. The would-wouldn't distinction implies a kind of acquaintanceship with the subject of judgements of plausibility and implausibility. It need not be acquaintanceship with an individual. Types of individuals also satisfy the distinction, as with the judgement 'A Manchester United supporter wouldn't root for Leeds.' Judgements of what a subject or type of subject would or would not do resemble, and may be a subject of, actions which are or are not *characteristic* of the subject or type of subject. It is not out of character of Harry to swallow the crazy politics of his girlfriend. It is out of character for football fans to be casual with their affections and their allegiances. Before their appreciation by the applied mathematics of gaming, the idioms of 'likelihood' and 'unlikelihood' fitted the would–wouldn't usage like a glove. 'Harry wouldn't do such a thing' courts equivalence with 'It's not like Harry to do such a thing', as does 'It's the sort of thing Harry would do' with 'How like Harry to do this sort of thing.'

Judgements of what is in or out of an agent's character resemble generic judgements. Judgements of what it is and isn't characteristic for a subject to do need not but can

be rooted in small samples of what that subject has and has not done to date. Of more central importance is the generality of such judgements. They are judgements in the form,'Subjects of such and such a character (e.g. people like Harry) do (or don't do) such things.' The similarity to generic judgements such as 'Tigers are four-legged' speaks for itself. Would-wouldn't judgements are, or are based upon, robust rather than fragile generalizations, are defaults rather than demonstrated facts, and are defeasible.

It is plausible that an agent did or would do X if doing X is in character, i.e. if doing X is the sort of thing he (or they) would do. An explanation of an agent's doing X is plausible if given the *explanans*, the *explanandum* X is something that the agent in question would do. Where an *explanandum* is not an action but a state of affairs or an event, then an explanation is plausible when given the *explanans* it would be in character for that *explanandum* to have occurred. So applied, talk of what is and is not in character for Nature to do is metaphorical. But it does no harm when what is and is not characteristic for Nature to do is understood as what is and is not *nomic* in the requisite correlations.

As we are conceiving of them here, relevance and plausibility are effectors of major economies in an agent's ecosystem of cognition and decision. Relevance is a matter of what presents itself to an agent's consideration, and is a small subset of what could have been considered. Plausibility engages generalities that need not be rooted in large samples (which are expensive to collect and manage) and which are not automatically defeated by a true negative instance. Thus is avoided the high cost of fragile generalizations, generalizations which require either repair or successorhood on the strength of a single contrary instance.

We have been speaking about individuals, about agents at the low end of the hierarchy of agency. We do not proposes here to discuss the other levels of agency in any detail. Suffice it to say that an agent's place in the hierarchy coincides with his (or its) need for, hence deployment of, scare-resource compensation strategies. The more highly an agent is placed in the hierarchy the more it can afford the time and the information with which to transact its affairs. In this, it is useful to recall Mill's point to the effect that institutions rather than individuals are the embodiment of inductive logics. Much the same can be said for classical systems of deductive logic. A related strength of institutional or collective agencies is that, unlike individuals, they are relatively untroubled by complexity, given that such agencies command the requisite computational capacity. A certain type of game-theoretic approach to a coming battle may be well beyond the calculational reach of any given general, but comfortably in the analytical ambit of his country's Defence establishment.

It is also well to note that an agent's place in the hierarchy is not a one-off matter. Within limits, the sort of agent he is is the sort of agent he can afford to be, which in turn depends on what is currently or prospectfully on his agenda. If he is writing a book on abduction, he should take pains and he should take time. He should even be prepared to give up if there is a notable lack of progress. But if the same person notices the back-door open of his presumedly locked-up house, he has options to consider and actions to take right then and there.

We have been attempting in these pages to make something of the contrast between reasoning as a practical matter and reasoning in science. If by science, we mean the theoretical formulation of its truths, the contrast we propose is reflected in

the distinction between practical and theoretical agents, a distinction which makes positions in a hierarchy of agent-types partially ordered by the resources that agents command. Conceived of in this way, there is little short of teamwork and access to a big computer that a practical agent can do to enlarge his command of resources and thereby advance his place in the hierarchy. We say that there is little he can do, but not nothing. Here is an example, which flows from the creative power of individual agents. Despite the scarcity of time, information and computational capacity, practical agents are capable of highly significant theoretic achievements. Practical agents are adept at thinking up theories. This has something to do with **heuristics**. Heuristics we understand in Quine's way. (Quine [55, 98–99].) They are aids to the imagination. They help the theorist in thinking up his theories. It cannot be put in serious doubt that in the business of thinking up his theories, there are some things the theorist cannot do without, including his most confident and enduring convictions about principles he thinks the theory must honour. Even so, not every belief required by the theorist to conceptualize and organize his theory need itself be a theorem of the theory. A case in point is any scientific theory eligible as input to the Löwenheim-Skolem theorems. All such theories must be extensional. Yet for all kinds of purely extensional theories, there isn't the slightest chance of our being able to think them up in a purely extensional language. In such cases, the intensionality of the thinking-up language is indispensable; but it would be a mistake to import those indispensable intensionalities into any theory governed by the Löwenheim-Skolem theorems. The mistake is bad enough to qualify for a name. We call it the

> *Heuristic Fallacy*: Let **H** be a body of heuristics with respect to the construction of some theory $T$. Then if $P$ is a belief from **H** which is indispensable to the construction of $T$, then the unqualified inference that $T$ is incomplete unless it sanctions the derivation of $P$ is a fallacy.

If the theorist bides the Heuristic Fallacy, he will be reluctant to enshrine in his theory those things that restrained him in his thinking the theory up, including those things that define the practicality of his practical agency. If, for example, his theory is a logic or formal semantics or an exercise in econometrics, it is completely open — indeed likely — that the theorist will sanction procedures or algorithms which are canonical in the theory, but which he himself, their inventor, could never run.

In this, the theorist is met with the ticklish problem of simultaneous avoidance of the Heuristic Fallacy and fidelity to the project of constructing ideal models of *appropriate* (that is to say *approximate*) concurrence with actual human performance. It is a task more easily prescribed than executed. What then, shall we understand a logic to be? We develop a fuller answer in work underway. For the present, we shall sketch the main idea. Again, let abduction be our example. Interesting abductive systems have been developed by researchers in artificial intelligence, including developments in diagnostic problem-solving. These logics run into a significant difficulty. The search procedures sanctioned by these logics are intractable: that is, unperformable in polynomial time. Such procedures are too complex for an individual's real-time computation. The search problems which these procedures are designed to solve are NP-hard, which in the 'traditional threshold of intractability'. (Bylander et al. [7, p. 157].) If the problem in question 'is 'small', then exponential time might be fast enough' but if 'the problem is sufficiently large, then even $O(n^2)$ might be too slow'. (Bylander et al.

[7, p. 157].)[12] It bears on this matter that there exists a time-honoured distinction between a logic and its search procedures, which is a matter for the meta-logic. It is a distinction that permits us to claim that a logic can claim success on the basis of the adequacy of its proof procedures, and independently of whether they admit of economical searches. There is no doubt that there are conceptions of logic (see below) for which this is a just observation. For other conceptions (see also below), it is a problematic claim.

A second example pertains to systems of relevant logic which closely orbit the basic Anderson-Belnap system. It is difficult to extend these systems in a natural way. For example, the Gamma Rule fails for $R\#$, which is the system of Peano arithmetic got from $R$. (See Fine [19] and Freidman and Meyer [22].) The failure of the Gamma Rule in $R\#$ greatly encumbers $R\#$'s proof-finding capabilities. Worse are the computational problems affecting $R$ quite generally. It is well-known that $R$ is undecidable. If the distributivity axiom is deleted from $R$, we get the decidable system LR. But LR is a computational horror. Its decision problem is at best ESPACE-hard, hyperuneconomically solvable by individual agents.

We grant that

> ¿From a purely logical point of view, abduction is a syntactical action or a theory $\Delta$ and a goal $Q$, consistent with $\Delta$, in a logic $(\vdash, S_{\vdash})$, yielding some additional data $\Delta_B$, consistent with $\Delta$ (denoted by $\Delta_B = Abduce\,(\Delta, Q)$), such that $\Delta$, $\Delta_B \vdash Q$. That is we 'answer' the question of 'what do you need to consistently add to $\Delta$ to make it prove $Q$?' (Gabbay [27, p. 204].

On a standard interpretation of 'purely', there is no doubt that the system here sketched is a logic of abduction. Purity has its place, and we have no wish to disdain it here. But in our approach, the logic of abduction will be one which builds upon and refines this core notion.

We understand a logic to be a *formalized* **idealization** *of a type of agent*. Given the striking and essential differences exhibited by agents at different ranks in the hierarchy of agency, it is easy to see that a logic which does well for a given type of agent does badly for agents of a different type. There is a standing invitation for logicians to commit this mistake, and the history of logic is liberally dotted with its commission. The propensity to make this mistake whereas in an essential structural feature of what constitutes a logic. A logic is an idealization of certain sorts of real-life phenomena. By their very natures, idealizations misdescribe the behaviour of actual agents. This is to be tolerated when two conditions are met. One is that the actual behaviour of actual agents can defensibly be made out to *approximate* to the behaviour of the ideal agents of the logician's idealization. The other is the idealization's *facilitation* of the logician's discovery and demonstration of deep laws.

There are limits to how far the theorist's idealization can go. It is by now widely agreed that classical first order logic is an excessive idealization of the behaviour of individuals, of agents at the bottom of the hierarchy of agency. Of course from the point of view of descriptive adequacy, *all* theories of human performance go too far, because all idealizations are descriptively inadequate. This is not to say that anything goes, or that nothing does. We propose the following limitation rule.

---

[12] '$O(n^2)$' denotes the complexity (order class) of examining the square of a number of data items $n$, each presumed accessible at unit cost.

*Logic Limitation Rule*: A logic is inappropriate for actual agents of type $\tau$ (or actual agents of type $\tau$ in relation to a given agenda) to the extent to which factors which make for agency of type $\tau$ are indiscernible in the behaviour of the logic's ideal agents.

It is well to note in passing the availability of machine modelling to serve — or try to — the requirements of a theory of individual cognitive agency. The great success of Turing's model in AI notwithstanding, it is unlikely that this is the way to go. For one thing, 'Turing machine programming is about the least perspicuous style of defining algorithms that has ever been invented.' (van Benthem [1999], 37). An alternative kind of approach suggests itself. The Game-Theoretic approach has already achieved something of a beachhead in logical theory. There are logical games for semantic interpretation (e.g. Hintikka [33], Lorenzen [44], Lorenzen and Lorenz [43]); for dialogue logic (e.g. Barth and Krabble [1982], Carlson [1983], Walton and Krabble [64], Mackenzie [1990], Girle [1993], Woods and Walton [1989], Hintikka and Bachman [1991], and Gabbay and Woods [2000], among others); and for the comparison of models. (Ehrenfeucht [17] and Fraissé [20].)[13]

Notwithstanding the prominence of the game-theoretic orientation, it too is met with nasty intractability problems, especially in dialogue logic. Nor does the game-theoretic approach exclude any notion of computability, never mind the difficulties to date (see here Moore and Hobbs [1996]).

Before bringing this section to an end, we take note of three particular challenges which the theorist of practical reasoning must try to subdue. This is not the place for intended solutions. It suffices that the problems are clearly set out and well-motivated. They are what we shall call the Complexity Problem, the Consequence Problem, and the Approximation Problem.

### 2.1.1 The Complexity Problem

In a purely commonsense way, individual agents are unable to deal with matters when doing so exceeds the time that can be afforded and the agent's computational power. This last is a constraint on complexity, and complexity here is a *first-level operational* matter. It should not be confused with *metamathematical* complexity. A case in point, as we have just seen, is the intractability of the decision problem for the relevant system LR. It is a problem no less hard than ESPACE-hard — a computational horror, as we have said. If anything is obvious about individual agency, it is how adept human beings are at discerning irrelevant information. This is done massively by the structure of consciousness itself. But even within consciousness, most of what an agent is aware of is irrelevant to the given task at hand. The obviousness of this fact carries over to one of its most interesting consequences: Efficient and timely management of the relevant — irrelevant distinction is *not* too complex for the individual agent to provide. So, in particular, we must avoid the mistake of uncritically endowing metamathematical complexity with *operational* significance. This we take to be the moral of the reason-is-repair slogan, and the several canons of minimal rationality that trail along in its wash.

[13] A good survey of logical games is van Benthem [1988, 1993].

Relevant logic aside, we join with Harman and others in saying that classical first order logic is too complex for the likes of us. That is to say, if *our* rules of inference included *its* 'rules of inference', and if we ran those rules in the way that they were run in first order structures, then, apart from some trivial exceptions, we would lack the time and the computational heft to make inferences at all. We are also minded to agree with those who claim that nonmonotonic reasoning (to take just one example) is more efficient, more psychologically real than its monotonic vis-à-vis. In one sense, nonmonotonic reasoning is less complex. But as studies in AI make clear, nonmonotonic reasoning is also more complex. In fact, any logic that deviates from the standard extensional logics involves an increase in complexity. It is not just that such systems are metamathematically complex; running their programs also represents a jump in complexity. So a question presses. How can, e.g. non-monotonic logic be simpler to use for practical agents and yet more metamathematically complex than first order structures which are difficult (to say the least) for practical agents to use? A case in point is consistency-checking. Consider the default rule:

$$\alpha : \quad \frac{\beta}{\beta}$$

which we can read as 'deduce $\beta$ if in context $\alpha$, $\beta$ is consistent'. The requirement is computationally complex for a machine. But typically a practical agent just 'intuitively' checks at little or no cost.

The problem, then, is this: how can it be the case that in everyday operational terms, individual agents are more or less good at ranges of tasks for which complexity is no particular problem, and yet, as studied by logicians and computer scientists, it is precisely those tasks that carry a degree of complexity which, if it actually obtained, would paralyze the individual's thought and action?

We have already noted that consciousness is a radical suppressor of complexity, and that computer simulation to date of individual agency have been unable to operationalize the distinction between conscious and nonconscious systems. The result of this is that in all—simulations of cognitive performance, there is vastly more information involved than any individual can consciously take in. Correspondingly, the simulating mechanizations exhibit (and handle) levels of complexity which are provably beyond the reach, often by several powers, of any conscious agent.

This appears to leave us with two options, both of which are underdetermined by any available evidence. One is to retain these over-complex systems, these aggregations of informational glut, and to postulate that they apply to agents *pre-consciously.* Below the threshold of consciousness, human systems are devourers of information, which enables them to handle substantial levels of complexity. We might judge it reasonable to think of the human neurological system as organic realization of PDP-architecture — computer analogues of the brain's own neurological network structure — the computer descriptions of which would then be of approximately the right type.

The second option is more radical, but it is no more foreclosed on by the available evidence than the first alternative. In exercising this option we would simply refuse to accept that any going logic or any going computer simulation stands a chance of elucidating individual agency in a realistic way.

Either way, we see it as a matter of a urgency that logicians and computer scientists forge serious, substantial, and long term partnerships with the brain sciences.

Before leaving this matter, it is well to emphasize that intractable, and otherwise unrealistic, theories $T$ of agency are devised by practical agents using cognitive and creative resources which do not find their way into $T$, either at all or in a descriptively adequate way. To some extent, their exclusion is justified by the necessity to avoid the Heuristic Fallacy. Beyond that, the exclusions constitute an abduction problem for the theorist. What best explains the exclusion from a theory of cognitive competence of those very cognitive skills which the theorist draws upon in constructing his theory? Various conjectures can be considered. One is that the theorist has a general idea of, but lacks a sufficiently detailed and descriptively adequate command of how those resources are deployed in real life. So, he activates the general idea in his theoretical model. Another is that the theorist's agenda is in part normative. If so, then his task must include the specification of norms which real life agents may and do deviate from in practice. The theorist will also be aware that in the very idea of a performance-norm is the requirement that actual behaviour counts as disconforming only if it is made out to bear a certain resemblance to the norms it violates. Another way of saying this is that only behaviour that approximates to a norm can be characterized as violating it. Why, then, is there often such a huge gap between what the ideal model prescribes and what practical agents are actually capable of? Our answer is that theorists have not yet succeeded, even where the need to do so has been recognized, in formalizing an approximation relation adequate for this theoretical task.

Examination of the historical record of theory formation in the areas of human performance suggests that idealized models fail to capture the actual — performable or near-performable — behaviour of practical agents. If this historical observation is correct, it must quickly be supplemented by recognition of the fact that theories that fail in this way may be seen as more faithful models of non-practical agency, of agencies of types that occur higher up in the hierarchy of agents. Agents so positioned we have dubbed theoretical. Theoreticity, like practicality, is a matter of the agent's command of the requisite cognitive and other resources required for cognitive performance; hence, twice-over, a matter of degree. Computational capacity is a case in point. Individuals, i.e. practical agents, have comparatively little of it, and collectivities, i.e. theoretical agents, have comparatively lots of it. A theory of human performance whose ideal models embed a lot of computational fire-power may fail as a model of practical agency and yet succeed as a model of theoretical agency.

This allows us to re-frame an important question. Why is it that theorists who seek to formalize practical or individual agency so often end up building models that fail for such agents and yet succeed, or come closer to succeeding, as models of theoretical agency? Our abduction is that this is the best that such theorists know how to do, that in questing for models appropriate to one type of agency they succeed in finding models that do well (or better) for other types of agency, which in their turn only approximate to the originally targeted agency-type. Here we meet with a methodological principle of substantial provenance. We call it the *Can Do Principle*. In its most basic form, the Can Do Principle bids an investigator of a question $Q$ in a domain $D$ to invest his resources in answering questions $Q_1^*, \ldots, Q_n^*$ from domain $D^*$ when the following conditions appear to have been met. First, the investigator is adept at answering the $Q_i^*$; and second, he is prepared to attest that answering the $Q_i^*$ facilitates the answering of the initial question $Q$. There is nothing to dislike in investigative practice governed by the Can Do Principle, provided there is reason

to believe that what the theorist attests to is actually the case. But as the present situation in, for example, rational choice theory, probability theory and mathematical logic itself clearly indicates, the attendant attestations sometimes stand little serious chance of being true. So the theorist plugs away at what he is able to do rather than what he himself has set out as his primary task.

Neo-classical economics is an instructive case in point. As is widely known, the neoclassical theory replaced the law of diminishing marginal utilities with the law of diminishing marginal rates of substitution. With the additional 'simplification' that goods are infinitely divisible, the theory had direct access to the firepower of calculus and could be formulated mathematically. Thus, for significant ranges of problems, it is easier to do the mathematics than the economics, with an attendant skew as to what *counts* as economics.

In its justified forms the Can Do Principle represents a sensible diversion of investigative labour, together with an implied (and usually rough) rank ordering. The Principle is justified when the enquirer has adequate reason to think that his investment in 'off-topic' work will eventually conduce toward progress in his 'on-topic' programme. It matters that whether the Principle is indeed justified is often indiscernible before the fact. In the natural history of the use of the Principle, its subscription is often tentative and conjectural, turning on features which give to the methodology of the investigation underway its own abductive character.

### 2.1.2 The Approximation Problem.

It remains our view that a logic is a formal idealization of a logical agent. The Logic Limitation Rule bids the theorist not to make too free with his idealizations. If the logician's or the computer scientist's ideal model is to be seen as modelling what actual agents actually do, what happens in the ideal model must be recognizable as the sort of thing an actual agent could or might do, or actually does. This factor of recognizability we have tried to capture by the relation of approximation, which bears on our problem in two ways. In the first place, an ideal agent's behaviour, *IB*, is recognizable as the sort of thing, *RB*, an actual agent really does, or could or might do, just in case, or to the degree to which, *RB*ing is an approximation of *IB*ing. But secondly, a theory T which fails to model with appropriate approximation the behaviour of agents of type $\tau$, may succeed in modeling the behaviour of agents of higher or lower type $\tau^*$. Even though T fails the approximation requirement in relation to the actual — or performable — behaviour of $\tau$-agents, T may still provide valuable insights into the workings of $\tau$-behaviour if the agency-type $\tau^*$, which fits T's norms more comfortably, is itself an approximation of requisite closeness of $\tau$-agency. We take it as a condition on a satisfactory theory of approximation that it preserves the intuitive inequivalence of these two notions of approximation.

The concept of approximation is borrowed from the natural sciences. The physics of frictionless surfaces is a case in point. Frictionless surfaces are mathematically describable idealizations of the slipperiness of real life, of the pre-game ice of the rink at the local hockey arena. Though the surface of the ice is not frictionless, it approximates to that state. There are limits on what to count as an approximation. After three periods of play, the surface of the ice is a less good approximation of frictionlessness than in its pristine pre-game condition. But no one will seriously

suppose that #04 sandpaper is also an approximation of frictionless, only less good still.

The approximation problem for logicians and computer scientists is the problem of specifying the mix of similarities and differences admissible by what they are prepared to call approximations of ideal performance. It is a difficult question. We may say with some confidence that no ESPACE-hard regime can be considered to be in the counterdomain of any approximation relation on conscious individual agents. But we don't want to restrict approximations to things which such a being could do if he went into training and tried really hard.

When, in 1999, the material of this chapter was presented to a meeting of computer scientists and electrical engineers, a member of the group said something along the following lines: 'I like your characterization of individual agency. And I too see a logic as a formal idealization of a type of agent. But are you sure that you're going to be able to write rules for this sort of case? I really need to see your rules!' It is a good question, and a hard one. It brings into apposition both the approximation problem and the complexity problem. The complexity problem is in part the problem of how much complexity in an ideal performance qualifies as that to which an actual agent's behaviour bears the approximation relation. And the question, 'Can you write rules for individual agency?' subsumes the question—or a question tantamount to the question—whether it is possible to make computer models of what is information-theoretically and complexity-theoretically *distinctive of* individual conscious agency.

### 2.1.3 The Consequence Problem.

In its pure classical state, a logical system can be seen as giving a principled description of the consequence relation. Non-classical variations can be understood in turn as principled descriptions of alternatives or rival consequence relations. We have already remarked on the difficulty such an approach presents the agency view of logic, that is, any view of logic in which a logical system is a formal idealization of a type of agent. The problem is that consequence relations are specifiable by truth conditions, or by proof-theoretic constraints, independently of anything that might be true of any actual agent.

It is possible to improve upon this austerely truth conditional approach to logic, that is, to a logic of agency, by taking a logical system to be an ordered pair $\langle S, \mathrel{|\!\sim}\rangle$ of a designated consequence relation $\mathrel{|\!\sim}$ and a set of instructions for proving when the consequence relation obtains in a context. In the example at hand (derived from Gabbay 1995), $\mathrel{|\!\sim}$ is nonmonotonic consequence and $S$ is a proof theory purpose-built for its peculiarities. In commonsense terms, a logical system of this sort is a principled description of the conditions under which an agent can declare (or recognize) a logical consequence of a database. The condition of $S$ clearly enough adumbrates the idea of agency, and we can see in $S$ an attempt of sorts to inferentialize the consequence relation. This is something Aristotle attempted 2500 years ago. Syllogistic consequence is just classical consequence constrained in rather dramatic ways, in ways that make the theory of syllogisms the first ever relevant, intuitionistic, nonmonotonic, hyperconsistent logic, or some near thing. Aristotle's question was in effect this: Can we get a plausible theory of inference from constraints imposed on the consequence

relation? This is also a question for proponents of $\langle S, \mathrel{|\!\sim}\rangle$. Can we get a plausible formal idealization of an actual agent by softening the consequence relation and harnessing it to a purpose-built proof theory? Our answer is that it depends on the type of agent, and his (or its) rank in the hierarchy. But it also seems correct to say that the lower down we go the less plausible the $\langle S, \mathrel{|\!\sim}\rangle$ approach becomes. But we note in passing that the more a logic of agency imposes constraints on the consequence relation, or the more it supplements it with additional structure, the more we remove from centre stage what we have been calling the purely classical view, in which logic is basically a bunch of truth conditions on the consequence relation.

## *2.2 Truth conditions, rules and state conditions*

The mathematical turn in logic changed (for a while) the conception of what a logic could and needed to be. In Frege's hands, logic needed to be re-jigged and retrofitted in order to accommodate the burdens of a particular thesis in the epistemology of arithmetic. On Frege's conception of it (but not Russell's) logicism was the view that since arithmetic is reducible to logic and logic is analytic, so too is arithmetic analytic, *pace* Kant.[14] Nothing in Frege's logicist ambitions for the new logic required it to address, still less to elucidate, the strict deductive canons of human reasoning and argument. When logicism expired (it could not survive the Gödel incompleteness result), the new logic was dispossessed of its historic *rasion d'etre.* It is open to wonder why the new logic didn't likewise lapse. That it didn't is a striking feature of the intellectual history of the 20th century, and it is explained in part at least by the Can Do Principle. In the span of time from 1879 to 1931, logic had become a dazzlingly successful intellectual enterprise — a growth industry, so to speak. In historically unrigorous hands, the logic of Frege and his successors reverted to its ancient status as a theory of strict reasoning, with evidence perforce of the Can Do Principle liberally at work. The boom times in recursion theory, proof theory, model theory and set theory are explainable by the fact that this was work that people were able to do, and to do extremely well; and it was seen as work that facilitated the overarching goal of producing a comprehensive logic of deductive thinking. Among those who knew better, the new logic permitted at least as much because it was found to be intrinsically interesting as that it was possible to do it well; and the Can Do Principle delivered the goods for that intrinsic interest.

As it has developed, mathematical logic, in both classical and nonstandard variations, examines the properties of structures. Such structures were not of a type that could pass for models of cognitive systems, except at levels of abstraction that made them unconvincing simulations of the actual practice of individual cognitive agents. For the most part, investigators of those structures hadn't the slightest inclination to think of them as models of human cognitive processes. They were studied because they *could* be studied, and because they were thought to be intrinsically interesting — as is virtually any enterprise that offers promise of well-regarded, long-term employment, which was the state of play in mathematical logic for virtually all of the past century.

---

[14] There are reasons simple and complex as to why Frege's logicism can't have been the same as Russell's. The simplest of of these is that Frege wanted logicism to prove the analyticity of arithmetic, whereas for Russell the truths of arithmetic were synthetic. The more complex story is well-told in Irvine [36].

Against this background, two historically important developments stand out. One involved the rising fortunes of nonstandard systems within logic itself. The other was the brisk evolution of AI. The two developments converged on an ancient idea; indeed it is the original *raison d'etre* of logic itself. Thus some (though not all) of the nonstandard systems and most of the approaches to computer logic were motivated by the desire that logic be a seriously deliberate account, or part of an account, of how thinking can and should be done. In the hands of logicians, this *was* an attempt to convert mathematical structures into cognitive systems; and, as was the case with relevant logicians, this was done by imposing nonclassical constraints on classical rules and operations. In the hands of computer science this was done by writing programs that simulated actual human performance. And, here, too, this was largely a matter of constraining the classical algorithms.

As we have seen, both these attempts to recur to logic's original motivation have met with various difficulties. Chief among them has been the high computational costs, higher than in classical systems, of running their algorithms, executing their protocols and deploying their rules. The results we have cited on the play of information on consciousness suggests an unattractive dilemma for the new, user-friendly logics. Either the new logics cannot be run by beings like us, or they can be and perhaps are run, but not consciously.

Logic's historic connection with thinking has always been with conscious thinking. If our present dilemma is well-grounded, then we would seem to have it that logic cannot discharge its historic mission (which would be another explanation of why mathematical logic doesn't even try).

One dilemma leads to another. Either the new logics are bad theories of human thinking, or they are possibly good theories of human *unconscious* thinking. Apart form the difficulty of determining which of these is likely to be true, there is the further difficulty that — historical anti-psychologism aside — theories of subconscious cognition have never been thought of as *logic*. We are now in the precincts of tacit knowledge, in which psychology has had what seems to have been a near-monopoly. The further dilemma to which this gives rise, is dilemma about logical rules. If rules of logic are thought of as having something to do with how human beings actually think, then by and large they are too complex for conscious deployment. On the other hand, unconscious performance or tacit knowledge is a matter of certain things happening under the appropriate conditions and in the right order, but it is unsupportably personificationist to suppose that this is a matter of following rules (an inclination which seems unshakably embedded in contemporary computer science.) *Façons de parler* being what they are, we can readily enough reconceptualize such 'rules' as causally enabling regularities; but then all semblance of logic as a prescriptive discipline is lost. A further dilemma, then, has it that logic has rules which humans can't conform their (conscious) thinking to or except for some fairly trivial conscious exceptions, logicality cannot be a matter of following rules.

Recent work on analogical thinking, emphasizes that '... thinking by analogy is an implicit procedure applied to explicit representations' (Holyoak and Thagard [34, p. 21]) Accordingly the goal of analogic is 'to make explicit how that implicit procedure operates.' (idem.). Plainly this cannot mean that the goal of analogic is to make explicit the rules which the subject explicitly runs to make the procedure work. It means rather that the goal of analogic is to make those rules or procedures explicit to

the theorist. Even this is a trifle tendentious. The theorist will explicitly conjecture a procedure of a *type* that he thinks plausibly applied to analogical thinking. He will say, for example, that the analogizer is adept at seeing relevant connections. In so saying the theorist is nothing but right that the correctness of his observation needn't involve his giving an account of relevance or specifying the conditions under which analogizers are good at detecting it; to say nothing of rules which the analogizer expressly deploys.

Explicit knowledge tends to be accessible to consciousness, and is therefore readily verbalizable by beings who have acquired the ability to speak. 'Using explicit knowledge often requires noticeable mental effort, whereas using implicit knowledge is generally unconscious and relatively effortless.' (Holyoak and Thagard [34, p. 22]) Here, then, is a mistake to avoid. Thinking is often conscious. When it its, it often involves propositional representations. It is entirely helpful to have good theoretical accounts of propositions—of how they are represented, of their grammatical structures, of their intentionality, and of those various properties and relations, possession of which bears on issues such as these. But it is a mistake to suppose that all our interactions with things we're conscious of are likewise objects of our consciousness. In particular, even if it is true that propositional representations require consciousness, it does not follow, and is not the case, that manipulating such representations is necessarily conscious. Still less does it follow that the cognitive manipulation of items of which we are conscious is a matter of following rules.

Logic is abductive in ways deeper than comprehended by Russell's regressive method in mathematics. (Russell [57]) This is true of all logic, not just what passes for abduction. Thus the logician conjectures about what it takes or what kind of thing it takes to get certain things done in ways that comport with an agent's cognitive behaviour, and what little or much the agent is able to simplify by way of laconic comments on the passing scene (in a memorable turn of phrase of Donald Davidson).

Logic is a model of a logical agent. Agency operates at various levels, central to which is the distinction between

- the conscious and propositional
- the subconscious and prelinguistic.

Logic accordingly involves

- a principled description of propositional structures, emphasizing properties deemed relevant to the description and/or evaluation of cognitive tasks

and

- a body of inferences about what goes on 'down below', and how it might influence or be influenced by conditions that obtain 'up above', i.e. propositionally
- conceptual analyses or definitions of the key ideas involved in the above two accounts.

Here is a conception in which logic is an enterprise with significant limits. Beyond the ingenuity of the theorist, chief among these limitations is the theorist's inability to inspect what goes on down below, on how propositional structures are actually handled, even consciously so. Returning to the example of relevance, beings like us are adept at discounting and otherwise disengaging from irrelevant information. Some of the time,

therefore, the propositional structure that has popped into a human head will be the output of his or her irrelevance-evading devices. But it cannot simply be assumed that there are linguistically representable properties of propositional structures that answer directly to the fact that it is a *relevant* propositional structure; which is a lesson lost on certain self-styled relevant logicians. The difficulty of determining the interconnections between what goes on up above and what goes on down below is noticeably less so in the negative cases. Thus, if the theorist-logician conjectures that the irrelevance-evading cognitive agent is someone who runs the algorithms that solve the decision problem for LR, then he attributes to the agent computational capacities which it is known that he cannot begin to approximate. This leaves a good question. What, on the agent's behalf, are we to make of the 'rules of inference' proferred by the theorist of propositional relevance?

It should not be forgotten that those who conceive of logic as exclusively the examination of propositional structures, with an emphasis on selectively important properties and on operations under which those properties are closed, are well-positioned to save themselves all the grief presently under review, and then some. All the more so, once the the move is made from linguistic structures to mathematical structures of higher abstraction. So to restrict logic makes more of a claim on prudence than is strictly justified perhaps, but there can scarcely be a logician alive who is unaware of such temptations.

There remains the fact that not all logicians are so methodologically circumspect, or ruthless. The new logic is awash in claims that go too far, in conjectures that are too much to bear by any fair measure. A good part of their problem flows from the very conception of logic that they are drawn to. It is a conception that originates with Aristotle. Aristotle wanted a comprehensive theory of argument. Owing no doubt to the ambiguity of the Greek word *syllogismos*, which our own word 'deduction' also inherits, Aristotle thought that a theory of syllogistic argument would also be a theory of deductive thinking. Indispensable to both projects is a theory of propositional structures which Aristotle called syllogisms. Syllogisms in this core sense are neither psychological nor dialectical entities. A syllogism is simply a triple of propositions answering to certain truth conditions. On the other hand, arguments in the sense for which he wanted a comprehensive theory are dialectical structures held to certain standards which are representable as sets of rules. Inference, or deductive thinking, is a kind of psychological modality subject at the descriptive level to certain psychobiological state-conditions.

Two things of importance require attention. One is that truth conditions, dialectical rules and psychobiological state conditions are three different things. What is plausibly supposed of a propositional relation such as implication (for example, that it answers truth conditions in virtue of which it is monotonic), cannot be plausibly said either of the dialectical rules of real-life argument or of the psychobiological state conditions of real-life belief-revision. No rule of argument will put up with the limitless supplementation of a valid argument's premiss-set, and no conditions under which an agent deduces a belief from a database will induce him arbitrarily and repeatedly to augment that database in ways that leave him wholly uninterested in whether his belief in the conclusion would (or need) change. It is true that Aristotle thought that the truth conditions of his purely propositional logic could be modified in ways that enabled them to be more plausible simulators of dialectical rules of dispute and ar-

gumentation and the psychobiological constraints on belief revision. Even so — and here is a second point that calls for attention — a problem arises that Aristotle could not have been aware of. It is the vexation that flows from the fact that imposing constraints on truth conditions with a view to their serving as dialectical *rules* makes for computational complexity on a scale that is hardly less than daunting.

Something of this difficulty is reflected in the entrenched affection of logicians and, especially, computer scientists, for anthropomorphizing the causal modalities of electric circuitry, or pretending that algorithms are actually instructions to an entity capable of reading and complying with them, when in fact they are causal triggers and regulators of digitalizable electronic flows (phantom algorithms, as we might say). Such processes bear a resemblance to what we are calling psychobiological state conditions, but even here there is a danger of a considerable misconception. There is reason to believe that under certain circumstances, psychobiological states are regulators of conscious states in beings like us. There is not yet reason to believe that the electronic etiologies which drive computer simulations of human cognitive effort succeed in producing anything that might pass for consciousness.

If we allow that a logic is a principled description of a logical agent and that human logical agency plays itself out both consciously and unconsciously, we leave it comfortably open in principle that the algorithms of an electrical engineer's making might enjoy literal application in matters of subconscious logical agency, and that the complexity discouragements that would bedevil the conscious running of such algorithms well might evaporate when run unconsciously in suitably layered architectures of the PDP kind. Talk of rules, on the other hand, is best reserved for the conscious domain, where it must be responsive to its very high levels of information entropy.

The various issues make it desirable to revisit the Heuristic Fallacy, which is the mistake of supposing that every proposition necessary for the theorist to believe in order to think his theory up is a proposition which the theory itself must formally endorse. The sheer attractiveness of the fallacy is hard to over-estimate. There is an entrenched methodology in philosophy and the abstract sciences generally according to which the theorist's core 'intuitions' must be preserved by any subsequent theory. The general inadequacy of this assumption need not detain us here (but see again. (Woods [70, Ch 8.]) Even so, if the theorist is not permitted to lodge in his theory any of his pre-theoretical beliefs, it is difficult to see how theories are possible. So some propositions, whose omission to believe would cause the theorist to fail to think up his theory, must be admitted. But which?

The Heuristic Fallacy (or the prospect of it) is important in another way. It is a way that offers encouragement to the logician concerned with matters down below. Logic, we say, is intrinsically abductive. It is a theory of how logical agents behave. Some aspects of that behaviour are attended by consciousness and are open to propositional representation and the discipline of rules. In other respects, the agent is a stranger to his own cognitive endeavours. He has no more access to the operations of his subconscious structures than his next-door neighbour or the cognitive psychologist down the street. The encouragement offered the logician of matters down below is in strictness offered not so much by the converse of the Heuristic Fallacy but rather a variation of it, according to which it is a fallacy to suppose that the mechanisms at work down below are nothing but the devices that constitute the cognitive agent's bag of heuristic aids. If it is supposed that only what is propositionally representable and

consciously accessible is subject for a logician's theory, then all else that facilitates cognition would find itself relegated to the category of heuristics. But the supposition in question is unreasonable. It suggests uncritical affection for propositional structures and over-ready susceptibility to the Can Do Principle.

Logic is a theoretical description of a logical agent. We may take it as given that in his various undertakings, the logical agent sometimes operates consciously and propositionally, and sometimes not. This alone makes the theoretical story of how logical agents operate an abductive story. We may also take it that in its various undertakings, logical agency sometimes involves the manipulation of propositional relations — or at least is constrained by them; that sometimes it involves what Harman calls changes in view; and that sometimes it involves reacting to proposals in argumentatively appropriate ways. All in, the logical agent operates at two levels, conscious and tacit, and engages or is influenced by truth conditions on propositional structures, state conditions on belief structures, and sets of rules defined for various argumentative structures. There is in what it takes for an agent to qualify for the status of logical reasoner some significant variety in conditions and wherewithal, which it would be folly for the theorist not to be alert to and disposed toward with an appropriate descriptive discrimination; all the more so when *levels* of agency are admitted to the theorist's palette. The story of the reasoning agent will vary with the propositional relations he is contextually placed — and able — to take into account, with cognitive inducements to change his mind or to think in some sort of different way, and with dialectical provocations to deploy various strategies of argument. If our logical agent is an individual, it will face these various conditions and incitements with scarce resources, implicit in which are limits on what counts as smart or rational even. If the agent is a theoretical agent, then its command of problem-solving resources enlarges in ways that match the degree to which the agency qualifies as theoretical, and criteria of success and failure change accordingly.

## *2.3 Rules Redux*

The logic of an individual's logical agency is a principled account of various practices. Since these are practices which cut across the distinction between conscious and unconscious processes, they are taken as flowing from capacities an agent possesses either tacitly or expressly or in combination. Three things are involved in the discharge of these capacities. One is the agent's manipulation of truth conditions on propositional structures; another is the deployment of and reaction to rules for making and for evaluating arguments — rules attending the agent's case-making proclivities; and the third is responsiveness to the causal inducements at play in the fixation of belief and the further aspects of changes in view. Since this trio of capacities cuts across the divide between the tacit and express, they will play with differential force depending on the particular theatre of operation. So, for example, an individual may have a change of mind in one of two ways, and at either of two levels. His new state of mind may be something his psychobiological conditions — his state conditions — put him in; or he may have changed his own mind in consequence of a case-making encounter with an interlocutor (and it is necessary to note that ultimately this present 'or' is not that of exclusive alternation). It may be a likelier thing than not that changes of the first sort are more frequently tacit than changes of the second, but there is no

question here of perfect concurrence. Whether his mind was changed for him or he changed his own mind, recognition of propositional properties (e.g. consequence or consistency) may have been in play; but it is not invariable that this is so, and here, too, such recognitions can be tacit as well as express.

Notwithstanding the critical differences between and among truth conditions, rules and state conditions, the rules approach is an entrenched habit among logicians. It is one thing to rail against bad habits. It is another, and better, thing to try to make them not matter, that is, to accommodate them in ways that minimize their sting. We may take it, then, that the postulation of rules and the attribution of rules-behaviour to logical agents is something to tolerate when the following conditions are met.

> *First*, it is reasonable to attribute to the agent in question the wherewithal (possibly tacit) to be situated *as if* he, she or it had consciously followed the 'rule' in question.
>
> *Second*, attribution of the rule thus qualified passes the other relevant tests.

When these conditions are met, we are free to attribute to real-life individuals what might be called **virtual rules**. In the spirit of the first condition, we might attribute to an agent conformity to the rule, 'Be relevant', when it is reasonable to suppose that the agent has resources, whatever they are, which place him in a situation that he would have been in, or that closely approximate to such a situation, had he had the means to follow the rule literally and had he done so in fact. The second condition secures a purchase in, e.g. the conjecture that since real-life individuals tend to transact their quotidian affairs in timely ways, such agents possess the wherewithal to evade or otherwise discount masses of information irrelevant to the task at hand. Thus 'Be relevant' could be a rule which the logical theorist sees fit to impose as a rational norm on the cognitive effort of real-life individuals without it being the case that, except in the attenuated sense presently in view, there is any reason to postulate that any agent's irrelevance-evading behaviour is the result of following the rule to be relevant. Rule-talk in logic, therefore, is largely a *façon de parler*. Once the *façon* is properly understood, there is no harm in the *parler*, for most of the rules cited by a logician — even a *nouvelle vague* logician — are virtual rules.

## 3 Concluding Remarks

Critics may complain that we have shattered a useful distinction between logic and psychology, and that we have done so without a satisfactory rationale. It is true that we reject certain interpretations of the distinction between logic and psychology. We reject any interpretation of that distinction which implies the mutual exclusivity of its relata. In other words, we reject that version of anti-psychologism which says that logic has nothing to do with how we think.

Still, it is fair to ask us to say with clarity where we see the difference, such as may be, between logic and psychology. Where, in particular, is the divide between the new logic and theoretical cognitive psychology? Our answer to this is that much of what the new logic comprehends is theoretical cognitive science, including relevant parts of computer science and any other discipline that bears in a principled way on how cognitive processes operate. Beyond that the new logic subsumes what cognitive psychologists are not typically very good at, namely, deep theories of the consequence

relation on propositional structures, and other target properties definable for propositional structures. Another way of saying this is that the new logic subsumes the 'old' logic, the logic that has come to be known as mathematical. There are two main reasons for this. One is that cognitive agents make essential (though often tacit) use of the logical properties of propositional structures as they discharge their cognitive agendas. Cognitive agents are mindful to a degree of the consequences of what they currently hold; they are interested to a degree in maintaining consistency in their databases; they are sometimes disposed toward truth- preservation in their argumentative and inferential practices; and so on. This being so, it is well for the logician of cognitive agency to have at hand good theories of properties such as these. A second reason for the new logic's favorable disposition towards the 'old' logic is methodological. There is a lot of pluralism these days about logical consequence and the other logical properties of propositional structures. But this does not alter the fact that, all told, our theoretical grasp of these properties is deep and substantial. At the very beginning of logic's long history, Aristotle attempted a bold experiment. In one way or another, the history of logic since Aristotle has been an extension of or a resistence to that experiment. What Aristotle sought to do was to retrofit the consequence relation, which is strictly a relation on propositional structures and nothing else, so as to make it a serviceable notion in the theory of argument and reasoning. The net effect of Aristotle's conditions on the syllogism was a consequence relation that was linear (hence relevant) non-monotonic, hyperconsistent and intuitionistic (See here [69]). Classical logicians of the modern era knew that there was no good reason to encumber the very ideas of consequence with these constraints. But as Aristotle knew, and with him the relevant, non-monotonic, intuitionistic and dialethic logicians of today, constraints such as these are essential if the truth conditions on the consequence relation are to have any chance of serving as rules of real-life argument and as state-conditions on real-life belief revision and belief update.

New logicians are satisfied that classical consequence (in the modern sense) can't realistically be pressed to such ends. But the jury is still out on nonclassical consequence. So nonclassical logics in the tradition of modern mathematical logic remain part of the logic of cognitive agency, and constitute an important part of the difference between the new logic and theoretical cognitive science.

## References

[1] Solomon Asch. Studies of independence and conformity: i. a minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70, 1956.

[2] Robert Axelrod. *The Evolution of Cooperation*. New York: Basic Books, 1975.

[3] D.S. Axsom, S. Yates, and S. Chaiken. Audience response as a heuristic case in persuasion. *Journal of Personality and Social Psychology*, 53:30–40, 1987.

[4] E.M. Barth and E.C.W. Krabbe. *From Axiom to Dialogue*, Berlin and New York: deGruyter, 1982.

[5] George Boole. *An Investigation of the Laws of Thought on whioh are Founded the Mathematical Theories of Logic and Probabilities*. Cambridge: Macmillan and London: Walton and Maberly, 1854.

[6] W. Brewer and G. Nakamura. The nature and function of schemas. In *Handbook of Social Cognition*, pages 119–160. Hillsdale, NJ: Erlbaum, 1984.

[7] Tom Bylander *et al.* The computational complexity of abduction. In John R. Josephson and

Susan G. Josephson, editors. *Abductive Inference: Computation, Philosophy, Technology*. Cambridge: Cambridge University Press, 1994.

[8] Laurie Carlson. *Dialogue Games*, Dordrecht and Boston: Reidel, 1982.

[9] Gregory N. Carlson and Francis Jeffry Pelletier. *The Generic Book*. Chicago: Chicago University Press, 1995.

[10] Christopher Cherniak. *Minimal Rationality*. Cambridge: MA: MIT Press, 1986.

[11] C.A.J. Coady. *Testimony: A Philosophical Study*. Oxford: Oxford University Press, 1992.

[12] A. Collins and M. Quillian. Retrieval Time From Semantic Memory, *Journal of Verbal Learning and Verbal Behavior*, 8:240–247, 1969.

[13] Peter Danielson. *Artificial Morality: Virtual Robots for Virtual Games*. London and New York: Routledge, 1992.

[14] Donald Davidson. Actions, reasons and causes. *Journal of Philosophy*, 60:685–700, 1963. Reprinted in Donald Davidson, *Essays on Actions and Events*, Oxford: Clarendon Press 1980, 3–19.

[15] David Deutsch, Artur Ekert and Rosella Lupacchini. Machine, logic and quantum physics. Los Alamos National Laboratory Preprint, orXiv:math.HO/9911150, 19 November, 1999.

[16] A. H. Eagly and S. Chaiken. *The Psychology of Attitudes*. Fort Worth: Harcourt Brace Jovanovich, 1993.

[17] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundementa Mathematicae*, 49:129–141, 1961.

[18] John Elster. *Sour Grapes: Studies in the subversion of rationality*. Cambridge: Cambridge University Press, 1985.

[19] Kit Fine. Incompleteness for quantified relevance logics. In Jean Norman and Richard Sylvan, editors, *Directions in Relevant Logic*. Dordrecht: Kluwer, 1989.

[20] R. Fraisseé. Sur l'extension aup relations de quelques proprietes des ordres. *Annales Scientifiques de L'E'cole Normale Superieure*, 71:363–388, 1959.

[21] Gottlob Frege. *Begriffsschrift, A Formula Language, Modeled upon that of Arithmetic, for Pure Thought*. Cambridge, MA: Harvard University Press, 1879.

[22] Harvey Friedman and Robert K. Meyer. Can we implement relevant arithmetic? Technical Report Techinical Report TF-ARP-12/88, 1988.

[23] Dov M. Gabbay and John Woods. Non-cooperation in dialogue logic: Getting beyond the goody two-shoes model of argument. *Synthese*, 2000. To appear.

[24] Dov M. Gabbay and John Woods. Editorial. *Journal of Logic and Computation*, 10(1):1–2, 2000a.

[25] Dov M. Gabbay and John Woods. More on non-cooperation logic. In *Lecture Notes on Artificial Intelligence*. Berlin and New York: Springer-Verlag, 2000b.

[26] Dov M. Gabbay. Editorial. *Journal of Logic and Computation*, 1990.

[27] Dov M. Gabbay. *What is a Logical System?* Oxford: Oxford University Press, 1994.

[28] Dov M. Gabbay. *Labelled Deductive Systems*. Oxford: Oxford University Press, 1996.

[29] Roderic A. Girle. Dialogue and Entrenchment. In *Proceedings of the Sixth Florida Artificial Intelligence Research Symposium*, Fort Lauderdale, FL, pp. 185–189, 1993.

[30] C.L. Hamblin. *Fallacies*. London: Methuen, 1970.

[31] C.W. Harper. Relative age inference in paleontology. *Lethia*, 13:239–248, 1980.

[32] Olaf Helmer. *Looking Forward: A Guide to Futures Research*. Beverly Hills, CA: Sage Publications, 1983.

[33] Jaakko Hintikka. *Logic, Language Games and Information*. Oxford: Clarendon Press, 1973.

[34] Keith J. Holyoak and Paul Thagard. *Mental Leaps: Analogy and Creative Thought*. Cambridge, MA: MIT Press, 1994.

[35] M. Howe. *Introduction to Human Memory*, New York: Harper and Row, 1970.

[36] A.D. Irvine. Epistemic Logicism and Russell's Regressive Method. *Philosophical Studies*, 55:303–327, 1989.

[37] Sally Jackson. Fallacies and heuristics. In *Logic and Argumentation*, pages 101–114. Amsterdam: North-Holland, 1996.

[38] Scott Jacobs and Sally Jackson. Speech act structure in conversation: Rational aspects of pragmatic coherence. In Rober T. Craig and Karen Tracy, editors, *Conversational Coherence: Form, Structure, and Strategy*, pages 47–66. Newbury Park, CA: Sage, 1983.

[39] Scott Jacobs, M. Allen, Sally Jackson, and D. Petrel. Can ordinary arguers recognize a valid conclusion if it walks up and bites them in the butt? In J.R. Cox, M.O. Sillars, and G.B. Walker, editors, *Argument and Social Practice: Proceedings of the Fourth SCA/FA Conference on Argumentation*, pages 665–674. Annandale, VA: Speech Communication Association, 1985.

[40] P. Kitcher. Explanatory unification. *Philosophy of Science*, 48:507–531, 1981.

[41] Theo Kuipers. *From Instrumentalism to Constructive Realism: On Some Relations Between Confirmation, Empirical Progress, and Truth Approximation*. Dordrecht: Kluwer, 2000.

[42] P. Lindsay and D. Norman. *Human Information Processing*, New York: Academic Press, 1977.

[43] Paul Lorenzen and Kuno Lorenz. *Dialogische Logik*. Darmstadt: Wissenschaft-liche Buchgesellschaft, 1978.

[44] Paul Lorenzen. *Formal Logic*. Dordrecht-Holland:D.Reidel Publishing Company, 1965.

[45] J.D. Mackenzie. Four Dialogue Systems, *Studia Logica*, LXIX:567–583, 1990.

[46] Ken Manktelow. Hove, UK: Psychology Press, 1999.

[47] J. S. Mill. *A System of Logic: Ratiocinative and Inductive*, 1843.

[48] Marvin Minsky. Frame-system theory. In R.C. Schank and B.L. Nash-Webber, editors, *Theoretical Issues in Natural Language Processing*. 1975. preprints of a conference at MIT, June 1975. Reprinted in P.N. ???? -Laud and P.C. Wason (eds.), *Thinking: Readings in Cognitive Science*, pages 355–376. Cambridge: Cambridge University Press 1977.

[49] D. Moore and D. Hobbs, Computational Uses of Philosophical Dialogue Games, *Informal Logic*, 18:131–163, 1996.

[50] D.J. O'Keefe. *Persuasion: Theory and Research*. Thousand Oaks, CA: Sage, 1990.

[51] C.S. Peirce. *Collected Works*. Cambridge, Mass: Harvard University Press, 1931. Published again in 1958 in Charles Hartshorne, Paul Weiss, and Arthur Burks (eds.).

[52] R.E. Petty and J.T. Cacioppo. *Communication and Persuasion*. New York: Springer-Verlag, 1986.

[53] R.E. Petty, J.T. Cacioppo, and R. Goldman. Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, 41:847–855, 1981.

[54] P.M. Pietroski. *Causing Actions*. Oxford: Oxford University Press, 2000.

[55] W.V. Quine. *From Stimulus to Science*. Cambridge, MA: Harvard University Press, 1995.

[56] Eleanor Rosch. Principles of categorization. In Eleanor Rosch and B.B. Lloyd, editors, *Cognition and Categorization*, pages 27–48. Hillsdale, N.J.: Erlbaum, 1978.

[57] Bertrand Russell. The regressive method of discovering the premises of mathematics. 1907.

[58] Roger Schank and Robert Abelson. *Scripts Plans Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

[59] Gila Sher. Logic: A theory of the obvious. In *Proceedings of the 1999 Conference of the Society of Exact Philosophy. Logical Consequences: Rival Approaches*. 2000.

[60] Edward E. Smith and Douglas L. Medin. *Categories and Concepts*. Cambridge, MA: Harvard University Press, 1981.

[61] Johan van Benthem. *Exploring Logical Dynamics*. Stanford: CSLI Publications, 1996.

[62] Johan van Benthem. Resetting the bounds of logic. *European Review of Philosophy*, 4:21–44, 1999.

[63] Y. Venema. A crash course in arrow logic. In *Arrow Logic and Multi-Modal Logic*, pages 3–34. Stanford: CSLI Publications, 1996.

[64] Douglas Walton and E.C.W. Krabbe. *Commitment in Dialogue*. Albany, NY: SUNY Press, 1995.

[65] Stephen Wolfram. Computer softwear in science and mathematics. *Scientific American*, 251:188, September 1984.

[66] John Woods and Douglas Walton. *Fallacies: Selected Papers 1972-1982*. Berlin and New York: Foris de Gruyter, 1989.

[67] John Woods, Ralph H.Johnson, Dov M.Gabbay, and Hans Jurgen Ohlbach. Logic and the practical turn: Introductory remarks. In Dov M. Gabbay, Ralph H. Johnson, Hans Jurgen Ohlbach, and John Woods, editors, *Handbook of the Logic of Argument and Inference: The Turn Toward the Practical*. Amsterdam: North-Holland, 2001. To appear.

[68] John Woods. *Death of an Argument: Fallacies and other Distractions*. Newport Beach: Vale Press, 2000a.

[69] John Woods. *Aristotle's Earlier Logic*. Paris: Hermes Science Publications, 2000b.

[70] John Woods. John Locke. *Argumentation*, 2001b. To appear.

[71] John Woods. *Paradox and Paraconsistency: Conflict Resolution in the Abstract Sciences*. 2001d. To appear.

[72] R.S. Woodworth and S.B. Sells. An atmosphere effect in formal syllogistic-reasoning. *Journal of Experimental Psychology*, 18:451–460, 1935.

[73] Manfred Zimmermann. The nervous system and the context of information theory. In R.F. Schmidt and G. Thews, editors, *Human Physiology*. Berlin: Springer-Verlag, 2nd edition, 1989.

Received 19 December, 2000