TOWARD WISDOM IN PROCEDURAL REASONING: DBI, NOT BDI

This paper is part of the following report:

TITLE: International Conference on Integration of Knowledge Intensive Multi-Agent Systems. KIMAS '03: Modeling, Exploration, and Engineering Held in Cambridge, MA on 30 September-October 4, 2003

To order the complete compilation report, use: ADA441198

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:

ADP021346 thru ADP021468
Toward Wisdom in Procedural Reasoning: DBI, not BDI

Kirk A. Weigand, Air Force Research Laboratory, Information Directorate, AFRL/IFSD, 2241 Avionics Circle, Wright-Patterson AFB, Ohio, (937) 904-9164, Kirk.Weigand@wpafb.af.mil

Abstract—Belief, Desire, Intention (BDI) procedural reasoning agent systems are founded on philosophic presuppositions that may be reassessed to aid the construction of sapient (wise) agents. This paper proposes a fundamental shift in order from BDI to DBI such that desire precedes belief and intention. This re-ordering may improve the ability of agent systems to wisely utilize procedural reasoning. This re-ordering to DBI is driven by interdisciplinary integration including the process philosophy of Alfred North Whitehead and philosophy of embodiment of George Lakoff and Mark Johnson. Their philosophic shift aids the re-definition of desire as the perception of affordances in the world. This new degree of freedom in the operation of procedural reasoning enables the use of human-in-the-loop evaluation qualities of aesthetic, affective, and emotional value structures or cue saliency constellations. These affective symbolisms may support dynamic re-structuring necessary for wisely adapting procedures to unknowable futures.

1. INTRODUCTION

Wisdom is the glory of man, not ignorance; light, not darkness! [1]

And of wisdom is the regard of place and the utterance of discourse according to measure and state. And of wisdom is decision; for man should not accept whatsoever anyone sayeth. [2]

Our hope of building sapient or wise information systems may require a fundamental integration of human and information systems. The Belief, Desire, Intention (BDI) procedural reasoning agent system can be re-framed to make this shift from an information system with a human interface (with a programmer and/or user) to a human and information system. This requires an explication of the relationship between humans and their information systems. Wisdom, knowledge, information and data must also be explained, disentangled and inter-related. The process philosophy of Alfred North Whitehead [24] and the philosophy of embodiment of George Lakoff and Mark Johnson [16,17] suggest a novel integration of perception, consciousness, emotion and reasoning to yield an important re-ordering from BDI to a new order of Desire, Belief and Intention (DBI). This apparently subtle change in order goes to the heart of tacit assumptions about the procedural reasoning methodology. Foremost among these assumptions is the implicit notion that beliefs — conscious reasons — precede and thereby serve to define subsequent desires and intentions.

This preliminary re-ordering and redefinition of BDI to DBI is accomplished by a method of interdisciplinary integration. Through support from various disciplines, a preponderance of theoretical support and empirical evidence enables this inductive synthesis of a preliminary, new and interdisciplinary conceptualization of procedural reasoning theory. This interdisciplinary approach challenges presuppositions that at first seem intuitive, i.e. that thoughts (beliefs) generate emotions (desires), but this notion stems from a tacit, culturally-widespread premise from psychology that feeling is an epiphenomenon of rational thought. Significantly, this assumption implicates and impedes the development of sapient systems by placing feeling in a subservient rather than an orthogonal role with respect to conscious reasoning. Notice that procedural reasoning systems are not explicitly named reasoning and valuing systems, yet they depend fundamentally on the value structures implicitly ascribed by their authors, coders and users. These value structures manifest in the priority schema and organization of objects in the code. If wisdom requires dynamic evaluation to adapt to unknowable future states of the world as they arise, having explicit control over the re-structuring of the code may enable this re-evaluation. This paper argues for a fundamental re-thinking of the BDI paradigm toward sapient multi-agent DBI systems. By exploring tacit or reduced presuppositions, the emergent property of a wise system may be discovered.

To theorize the existence of a wise, multi-agent system, first this paper will take a pragmatic, selective look at the existing BDI paradigm to help reveal the tacit assumption that the emotional value state of the system is produced by the rule set describing beliefs. With this emancipation of a new degree of freedom in the design of multi-agent systems, an alternative, complementary and orthogonal relationship is established between emotionally-based value structures and rational beliefs. The orthogonality of emotional, affective value structures with respect to rational logic has been supported from cross-cultural anthropology, process philosophy, cross-cultural psychology, group dynamics and sociology [22], so this will not be covered in detail here. A
rework of all the semantics, propositions and lemmas from BDI to DBI may be a difficult job and is left for future work. With the plausibility of the orthogonality of affective value structures to rational logic supported philosophically, this new degree of freedom is used to posit a sapient, dynamic process as the continual realization of internal logic, objective context, affective evaluation and human-in-the-loop grounding from salient cues and expectations.

2. Why Re-order BDI?

Of course, this proposal to re-order BDI to DBI may meet many, significant objections. This new theory may re-open a host of solved problems, may be impractical or require extensive re-work before a mature system is available. A new BDI theory must elegantly solve not just an anomaly in the existing BDI approach, but some crisis. [15] This crisis is now apparent as we attempt to move beyond intelligent and knowledgeable systems toward wise systems. The sapient system must not only deliver – as in a knowledge system – the more relevant data given the context of the system and user(s), but also employ a causally efficacious induction that rolls-up this relevant data into information that matches the expectations of subject matter experts and that guides the untoward expectations of the typical novice. For example, a wise human is a subject matter expert who chooses the appropriate cues to satisfactorily and sufficiently (to satisfice per Herbert Simon) recommend a causally efficacious course of action. The causal efficacy of this expert’s decision produces a fortunate and satisfactory effect as grounded by reality. [24] Conversely, an expert or novice can misread cues or mistakenly focus on irrelevant stimuli and, consequently, produce an erroneous or ill- advised decision with an increased probability of an unfortunate outcome. Whether an outcome is fortunate or not depends on how fortune is defined and evaluated. So, a means for this evaluation becomes central to the crisis in the ability of multi-agent systems to consistently present the expected cues to wisely describe a dynamically unfolding situation.

BDI contains a means for evaluation of causal efficacy, but the evaluation schema appears to be buried in the code and therefore not accessible for adapting to unpredictable contexts. Systems exist that attempt to resolve contradictions through rule-based schemes, but they have yet to exploit emotions as a means of evaluation. If the sapient system is viewed as a cognitive prosthesis [10] for causally efficacious decisions, the sapient system should enhance the sensitivity to what is key to making good decisions. Historically, what is good and appropriate was the sole purview of the human. The human chose the appropriate application for the given context, but with the advent of sapient systems, the system must dynamically assess goodness, relevance, and appropriateness toward the accomplishment of some efficacious purpose across diverse situations. This tie to prosthetic functional enhancement and affordances fits with the intentional stance of Daniel Dennett, where BDI attributes are ascribed to predict behavior [18]. The sapient system may be able to find the appropriate intentional stance, if the evaluation schema can be accessible as a kind data used to guide the structure and prioritization of the code. The novel introduction of a cue saliency structure [21] as an affective constellation of values [6,23] may provide a sapient system with this other degree of freedom to adjust and restructure the priority and attention of the agents in the system. If these agents can work sympathetically [22] in pursuit of the satisfaction of causal efficacy and adjust their priorities accordingly, the system may be said to be wise.

Some may argue that the order of BDI versus DBI is somewhat arbitrary in their actual applications. If true, this may indicate the trivial role of desires in the current BDI implementations where desires are simply second-order beliefs. Rao and Georgeff’s application of BDI and naming as a procedural reasoning system suggests a subservient role for desire. In their applications, desires appear to be abstractions of beliefs. At the heart of BDI is Michael Bratman’s construction of Gricean creatures via thought experiments about simple imaginary creatures. [3] Bratman begins with Gricean Creature 1 – an imaginary creature with beliefs but driven incessantly by desire. From Creature 1, a more robust Creature 2 is imagined with the added complexity of considered desires. Considered desires are desires as accorded with beliefs. This suggests an introspective agent that attempts to integrate its desires with its beliefs. Here, Bratman leaves open the independence of desires from beliefs. Rao and Georgeff seem to have taken considered (rational) desires as the definition of desires, thus setting the order in BDI. As discussed below, interdisciplinary work supports a revolutionary notion that desires are a non-rational and independent dimension of human awareness. Lakoff and Johnson’s extensive work supports this thesis. [17]

If the order of BDI is critical to its architecture and application, then this leads to an impasse as a wise agent system. If the system can only reason in an analytic deductive sense, it can never induce other interpretations of the data that may have other potentially valuable kinds of causal efficacy. If BDI is attempting to wisely reason inductively as well as deductively, it must explicitly deal with other value structures that are not in accord with its current belief set. Different criteria may result in different fortunate evaluations of a phenomenon. But if BDI’s considered desires are derived from and consistent with its current set of beliefs, it logically cannot propose desires that do not accord with its current beliefs. Yet, if its current beliefs turn out to be unfortunate (not causally efficacious by the current evaluation scheme), the system is stuck with no way to change its evaluation criteria. This conclusion is supported by the 1992 call by Georgeff for significant advances in dynamic machine interpretation for intelligent decision systems, yet he describes this as a need to improve the process of analysis and reasoning rather than synthesis and valuing. [8] Therefore, if the order of BDI is fixed, this
may imply a philosophic dead end for BDI as a sapient system. By positing the independence of desire from belief and from considered desire, this new DBI approach suggests that desire is a complement and an equal partner to belief.

3. DBI as a Human and Information System
Knowledge as Inherently Human

Implicit so far in this discussion of the re-ordering of BDI to DBI is the debatable attempt to mimic human consciousness and awareness in the agent system, yet since wisdom appears to be deeply entangled in human conscious and unconscious avenues of awareness — including tacit knowledge and intuition — this DBI agent framework needs to have tight coupling to humans to even begin to approximate wisdom. This tight coupling may be achieved by mimicking our current, ever-expanding understanding of the human mind as embodied in the brain and the rest of the body. The explicit premise of this re-ordering is that a sapient DBI system must be a human-in-the-loop integrated system. Some of the attraction to BDI may be the ease with which designers and programmers can translate notions of beliefs, desires and intentions and map them to real-world applications. In this way BDI provides some proficiency for including humans-in-the-loop a priori to establish the appropriate value structures during design and coding, which results in subsequent, run-time evaluation by the multi-agent application. In DBI, this may be enhanced by run-time, human-in-the-loop, and dynamic re-evaluation with the goal of adaptive re-structurization of the code aided by human expectations through in-situ knowledge elicitation.

Fasli notes that BDI is useful because of its flexibility to model different degrees of realism. [7] Once an application is built in BDI, the degree of realism becomes fixed in the implementation. As an advance over this fixed realism, a new sapient DBI system might dynamically determine which kind of realism is better suited and wiser to employ given a dynamic context.

A sapient system, from this perspective, requires humans in the loop to establish knowledge. McQuay's review of tacit knowledge research finds that knowledge does not seem to exist except in the mind of humans. [19] Information and databases may provide elements of knowledge, but without coupling to human purpose and context, the relevance of this information cannot be ascertained, organized, presented and used. For example, a model of the physical structure of a chair might at first appear to be knowledge, yet unless some agent cares about the structural affordances of chairs, this may be extraneous information, e.g. the material composition of the chair does not directly contribute to the agents functional needs for the chair to support weight. Some additional information must be integrated with this information to yield usefulness via load analysis data. A human may "know" that chairs made of wood are strong enough to sit on. This integration of humans with computer agents leads to the discussions of autonomy and agency that are beyond the scope of this paper, yet as an explicit premise, humans are ultimately responsible for any agency granted to software agents. Therefore, the evaluation criteria used in sapient systems must entail human tacit knowledge and intuitive understanding. This dictates a close coupling and coordination between the human and information system, which is supported by Woods' claim that more complex autonomous operation demands higher coordination. [25]

The Crisis in BDI: the wake that steers the boat

Below, the metaphor of a boat and its wake is used to demonstrate the crisis in the BDI framework if it is to be used as a sapient system. In the attempt to move BDI from a procedural reasoning system to one that chooses procedures wisely, a logical circularity becomes apparent, as the structure for evaluation requires it to question its own structure. By what meta-structure can the system evaluate its evaluation? A re-ordering of BDI to DBI is much more than a change of workflow in the code. Instead, DBI seeks to tie the human intuitive genius of evaluation with the computational brilliance of traditional procedural reasoning algorithms and processes.

Where BDI has gone astray of being a human analog is in its linkage of sensing to belief, because cognitive psychology suggests that sensing is more directly driven by desire, but cognitive psychology appears conflicted because it also says that desire results from belief. BDI associates belief with sensing the environment. [12] These founders of BDI built knowledge areas (KAs) out of beliefs, goals and intentions, but their notion of desires as goals seems to reduce desires to behavioral objectives, which are difficult to distinguish from intentions. Fasli suggests from Searle that pro-attitudes such as desires and wishes have a "world-to-mind" mapping. [7] This supports a tie between desires and sensation, which in turn suggests a tie to an aesthetic sensitivity to the world. Intuitively, we seem to feel the world before we are able to contemplate it. Unfortunately, BDI's assumption of three cognitive states — information, motivation, and deliberation — associates desire with the motivational state as a pro-attitude, which disconnects it from the "world-to-mind" sensory inputs and places belief as the information gathering intermediary between the world and motivation. [7] Below the cognitive psychology research of Lakoff and others is used to resolve the conflict in the primacy of belief and desire.

The perceptual sub-system in BDI is linked to the beliefs about the world and this appears to be consistent with Carbo, Morlina and Davia's definition of beliefs as assumptions about the world as perceived. [4] However, this is not supported by cognitive psychology studies of naturalistic decision making of experts in complex situations. For example Klein's Recognition-Primed
Decision (RPD) model has empirically verified that expert's subconscious perceptions of cues determines their attention and expectancies. [13] Conscious beliefs can arguably set the stage for this subconscious experience and in this way past beliefs seem to precede this experience of current desires, which is the basis of psychotherapies such as Rational Emotive Therapy. [11] This argument, while apparently true, becomes mute when we define desires as the emotional experience of reality and considered desires as the rational introspection over current desires in the wake of past beliefs and intentions.

The metaphor of a boat plowing the water of experience may help show how evaluation and perception bring the unknown world into consciousness. The prow is emotional subconscious awareness (desire). Its attitude on the world is determined by beliefs represented by the current settings of throttle and rudder in the cockpit. This nautical trim or balancing of the boat establishes its attack through the water. Intentions are represented by the actual thrust and vector of the propeller and rudder. Intentions like the rudder make contact with the water of reality at the stern. In this metaphor, conscious awareness grows from fore to aft, from bow to stern. As the prow's direction is established by the history of interactions of the rudder and propeller with the water, so too are present desires vectored by past conscious beliefs and intentions. Still at any moment the boat is led by the prow of subconscious expectations as set by past states of the system. This metaphor implies an on-going cycle of feeling, thinking and acting as supported by the process philosophy of Alfred N. Whitehead. [24] Lakoff empirically validates this aspect of Whitehead's speculative philosophy via cross-cultural linguistic research. [16]

If this metaphorical boat experiences the world in a changing context according to Klein's RPD model, it can only respond to the world at that instant by its current configuration, speed, orientation and trim. The expert, according to Klein, seems to perceive what is relevant while excluding what is irrelevant and thereby rapidly converges on a causally efficacious perception of reality. Initially, all the expert can do is set the speed, trim and attitude of her perceptual system. For example, a firefighter can only be "ready" as she steps through the door of a burning building. All the years of training and experience boil down to that single step. Her prior beliefs, intentions and rationally considered desires shape that readiness, yet as she advances into the unknowable reality of a burning building, her perceptual state aesthetically values the plethora of sensory inputs in a way that maximizes survival and goal accomplishment – her definition of causal efficacy. As the metaphorical prow of the boat of awareness moves into the unknown, the expert pilot has trimmed her craft to just the right cues to steer the boat through unknowable water ahead. While the wake of awareness enables reflection, introspection and learning, it can only serve to set the stage for future perception of the unknown.

The current definition of desire in BDI may still be tenable, however its institution as lemmas, semantics and propositions may require significant changes. The above interpretation of desire as jointly an aesthetic sensitivity to causally efficacious cues and as subconscious emotional attitude toward the unknown world, may be viewed as the more general case of the more specific definition of desire now in use in BDI. Currently, Carbo, Molina and Davila define desire in BDI as preferences over future states of the world. [4] Preferences can be thought of as manifested in the current trim of our metaphorical boat and desires can be seen as sub-conscious embodiment of preferred reality. [17] Cue saliency constellations may then be said to be the perceptual, aesthetic sensitivity that guides attention toward salient patterns in the world. What is salient at any given instant is dependent on the last vector of purpose. Our firefighter must have the right attitude to survive the burning building.

To summarize this claim, desire is not only a preference for future states, but is more generally a cue expectancy constellation of valuations that manifest as perceptual states of attention to certain cue patterns in the environment. These patterns may be physical low-level sensory perceptions from the senses or multi-modal, highly abstract patterns. Regardless of the level of abstraction, these patterns are felt as emotional states and as an aesthetic sense of the world that is a subconscious prehension preceding conscious awareness. [24] At the instant of belief or disbelief, this prehension becomes available for rational understanding and subsequently drives intention as the rudder of the boat. Intention then can be defined as a process that is the outcome of subconscious cue expectancy constellations aesthetically filtering external states of the world according to a goal of survival in the world as constrained against the history of beliefs about that world. The pilot trims the attitude of the craft. Introspection about this attitude of the intentional state may be defined as the current role of the agent. Introspection about past beliefs in the wake of current beliefs defines learning. Introspection about the current cue expectancy constellation defines an awareness of the emotional state of the agent. Regardless of whether or not the agent introspects on its current emotional state and records it as a belief (considered desire) for future use, it still must have an emotional state. An agent that is never introspective about its current cue expectancy constellation, is oblivious of its emotional state, yet is nonetheless beholden to aesthetic perception by it.

**Whitehead's Process Philosophy**

Process philosophy supports a re-ordering of BDI to DBI as well as the subsequent re-definition of desire in emotional, affective, subjective and vaguely associative terms such as cue saliency constellations. In the search for sapiet wise systems Whitehead provides a bridging philosophy between the logical and the emotional as an advance beyond Kantian dualism. Wise systems need to judge well. According to
Whitehead, “Judgment is the decision admitting a proposition into intellectual belief.” [24] “A proposition is an element in the objective lure proposed for feeling, and when admitted into feeling it constitutes what is felt.” [24] This may be equivalent to desire being considered and thereby becoming a considered desire. “The word ‘decision’ does not here imply conscious judgment, though in some ‘decisions’ consciousness will be a factor. The word is used in its root sense of ‘cutting off.’” [24] A feel, think and act cycle cuts off the world as it is perceived and judges it as a decision in that moment. Carriers of intelligence may be equivalent to Whitehead’s propositions as considered desires, and therefore, judgment is the admission into intellectual belief of these carriers of intelligence.

From a Whiteheadian perspective, desire, appetition or more specifically, prehension is the most primitive element of unconscious experience by which higher abstraction beliefs and intentions are produced. If desire becomes the primary component in a DBI framework, sensory input becomes an aesthetic and felt experience that creates patterns of relevance. “The primitive form of physical experience is emotional – blind emotion … [in other words this] primitive element is sympathy, that is, feeling the feeling in another and feeling conformally with another. [24] This primitive pre-conscious kind of sympathy describes an aesthetic sense by which impinging sensory signals are converted into contrasts of pattern. Relevance, thereby, establishes feelings of attraction or repulsion to various cues and patterns in the world while ignoring less useful patterns.

This kind of primary shaping of sensed patterns suggests that desires may fundamentally constrain subsequent adaptation of beliefs by shielding evidence contrary to desire, perhaps, in spite of beliefs. Here, seeing is believing, unless higher abstractions of belief intervene in subsequent feel-think-act cycles to feel differently and, henceforth, see the world anew. Intentions follow as a satisfactory decision – in the moment – according to congruence of desire and belief.

Sympathetic Communication of Desire via Intent

If sympathy is considered a primitive aesthetic feeling of desire in a DBI conceptual model, then intent is the product of an aesthetic feeling of an entity for other entities, which defines the relevance of contextual data by which beliefs are confirmed and actions intended. A sympathetic communicative action model [9] suggests that a third channel of media richness is used in human communication to communicate the aesthetic sense of what is important or meaningful during speech actions. This third channel of “sincerity” supplements logical content and normative role-based aspects of speech. According to Habermas speech acts can fail if any of these three channels of information are not delivered. A speech act can fail if it doesn’t make logical sense, if it is inappropriate according to the role of the speaker and if the speaker is not sincere. This third dimension of sincerity relates to a proper sympathetic mapping from speaker to listener. [23] Behavioral components of speech such as tone, inflection and facial gestures, describe an affective or emotional dimension to speech that adds value structures to the logic and illocutionary aspects of the spoken language. This affective behavior appears to tap the primitive aesthetic sense of relevance and maps the previous intent to communicate the speaker’s ideas to the next cycle of desire, belief and intent in sympathy with the listener.

Cue Saliency Constellations

The definition of desire as a complement to belief accords with Modeling Field Theory’s neural architecture [20] where a fuzzy association subsystem works with a fuzzy object modeling subsystem to jointly and dynamically adapt to the changing world and converge on an object model and a conceptual model that are mutually constrained. In the language of the proposed DBI theory, desire, and more generally, a cue expectancy constellation, form the association subsystem of MFT neural architecture while belief represents its object modeling subsystem. Desire drives perception toward relevant cues, which expects certain objects with certain purposes. As objects are seen they confirm the associative frame of reference. A chair becomes a chair in conscious awareness in the context of an affordance for sit-ability, if you intend to sit on it. Cazeaux makes the tie between the ecological theory of perception via affordances of Gibson, Maurice Merleau-Ponty’s body schema and Lakoff and Johnson’s embodiment. [5] By way of Cazeaux, wisdom in humans must be embodiment as causally efficacious metaphors. Therefore, cue saliency constellations may be posited to be the symbolization representing metaphor as embodied in a human. Given the link of desire and emotion to this embodiment, cue saliency constellations may well be affective symbolic representations of metaphor encoded in the human perceptual system. This has fascinating implications for the design of an integrated sensor, information and human system.

4. CONCLUSION

The principal result of this work is the potential plausibility of the notion that BDI may be adaptable for use as sapient systems. This may offer the advantage of leveraging on the vast effort already expended on this system. The main disadvantage is that the re-ordering of BDI to Desire, Belief and Intent (DBI) may invalidate some of this work. The advantage of DBI is the new degree of freedom from the disentanglement of emotional value structures as desires from the logical reasons as belief sets. Further research in this area may include a formalization of DBI semantics, propositions and lemmas. This would enable a comparative study of the strengths and weaknesses between the two approaches. Another avenue of future research is linking
cue saliency constellations to the human as an *in situ* knowledge elicitation human-computer interface. The symbolization needed to elicit cue saliency constellations might help define the communication necessary for sympathetic agents to wisely share scarce resources including the scarce attention of humans in the loop.

5. REFERENCES


