Autonomous Systems Behaviour

Derek Hitchins

Introduction

Autonomous machines have fascinated people for millennia. The Ancient Greeks told of Talos, the Man of Bronze, whose supposedly mythical existence has been given unexpected credence by the recent discovery of the Antikythera Mechanism. Could the Greeks really have created a giant mechanical man to protect Europa in Crete against pirates and invaders? Long before that, the Ancient Egyptians created the Great Pyramid of Khufu as an autonomous psychic machine to project the soul *(ka)* of the King to the heavens where he was charged with negotiating good Inundations with the heavenly gods.

In more recent times the sentient machine has been represented as logical, rational yet invariably inhuman. HAL[1] from Stanley Kubrick's *2001: A Space Odyssey* is an archetypal example, with HAL being dedicated to completing the mission regardless of the crew. Before that, in 1942, Asimov invented the Laws of Robotics, later made famous in books and the film *I Robot:*[2]

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey any orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

These seemingly fool proof laws encouraged the robots to imprison the human population, supposedly to protect humans from self-harm, consistent with the first law.

It seems to be the case that we humans generally expect automatic machines to be flawed, and that they will become so intelligent yet inhuman that they will take over the world: an uncomfortable juxtaposition of incompetence and omnipotence…

Publicised advances are being made in Japan, where they have robotic figures that can run, climb stairs, and some even display facial expressions. Honda's Asimo is able to learn about, and categorize objects. Nursing robots, and others, are being made rather smaller in stature than their equivalent human: Japanese designers seem well aware of the innate fear reaction that autonomous machines arouse in many humans.

One possible way forward, which may enable us to create smart, autonomous machines without the associated fear, might be to make the machines more human-like, so that they could act and interact with people on a more comfortable basis. Could we, perhaps, make machines that behaved in some respects like people, making acceptably ethical decisions and moral judgments, even exhibiting sympathy and empathy with victims and patients?

---

1 The name HAL derives from IBM, with each letter being one before its equivalent in the alphabet…H is one letter before I, and so on.

2 "I Robot" is a play on Descarte's *Cogito Ergo Sum* – "I think, therefore I am:" proof of existence.

Even if we could do these things, would it be *wise* to do so? Or, would we, in effect, be creating beings with similar underlying flaws to Asimov's infamous Robots?

Meeting the Challenge

Complexity

Any autonomous machine capable of establishing its own purpose, making plans to achieve that purpose and executing those plans is inevitably going to be complex. We can imagine that such machines would have some form of 'brain' or perhaps a central decision-making and control centre. Moreover, for a machine to emulate a human in terms of physical dexterity suggests that the machine will inevitably be complex – not as complicated, perhaps, as a human, but nonetheless complicated.

Nature has much to tell us about complexity, although her lessons may be hard to learn. Life itself appears to be related to complexity. Up until some 580 million years ago, life on Earth was confined to simple cells and cell-colonies. Over the next 70-80 million years there was a mass diversification of complex life forms, the so-called Permian explosion, followed by periodic extinctions, ending in the 'great dying,' which marked the end of the Permian Period and the start of the Triassic.

Complexity, in Nature at least, appears to 'auto-generate,' and then to collapse before starting up again. Human civilizations follow the same pattern, with the same collapse: business organizations likewise. This phenomenon suggests underlying principles:

1. Nature creates wholes – organized, unified wholes – whence *holism*…
2. Wholes comprise interacting organic subsystems – whence *organicism*
3. Organic subsystems interact synergistically within their context and environment – whence *emergence*
4. Man and Nature synthesize unified wholes, with emergent properties, from complex, interacting subsystems, themselves wholes… - whence *hierarchy*
5. Complexity creates both order *and* disorder…
    a. The Big Bang created stellar systems; galaxies; clusters; super-clusters… and destructive Black Holes
    b. *Hymenoptera* (social insects) and *termites* create hives, colonies, bivouacs, etc., which eventually collapse only to be replaced by newer versions
    c. *Homo sapiens* create families, societies, cultures, civilizations… that eventually collapse, only to rise again in revised form.

Consideration of this ubiquitous pattern of synthesis and collapse suggests a *Law of Complexity [1]:*

> *Open, interacting systems' entropy cycles continually at rates and levels determined by available energy*

This stands in counterpoint to Kelvin's Third Law of Thermodynamics, which proposes that the entropy in a closed system will either remain the same, or will increase.

The tentative Law of Complexity also suggests that homeostasis (dynamic equilibrium) in a particularly complex body may be problematic.

Complex Growth

Figure 1 arranges some of these interrelated ideas into a hierarchical structure. At hierarchy level N-2 there are five complex systems (ovals), each containing six complex interacting subsystems. Since the subsystems and the systems are all open, they will all adapt to each other as they exchange energy, material and information. The result of action, interaction and adaptation is the emergence of properties of the whole (lowest layer) as shown at the right.

Hierarchy level N-1 is shown as also having five system groups. Each of these five will comprise the systems of Level N-2, but only one of these contributors is shown to prevent the figure from overflowing. Similarly, all of the complex systems at Level N-1 contribute to/synthesize only one of the smaller ovals in the six oval figures at Level N-0, the nominal level. The potential for complexity is already phenomenal, as each open element throughout the network/hierarchy adapts – inevitably, the whole would behave dynamically in simulation as well as reality.

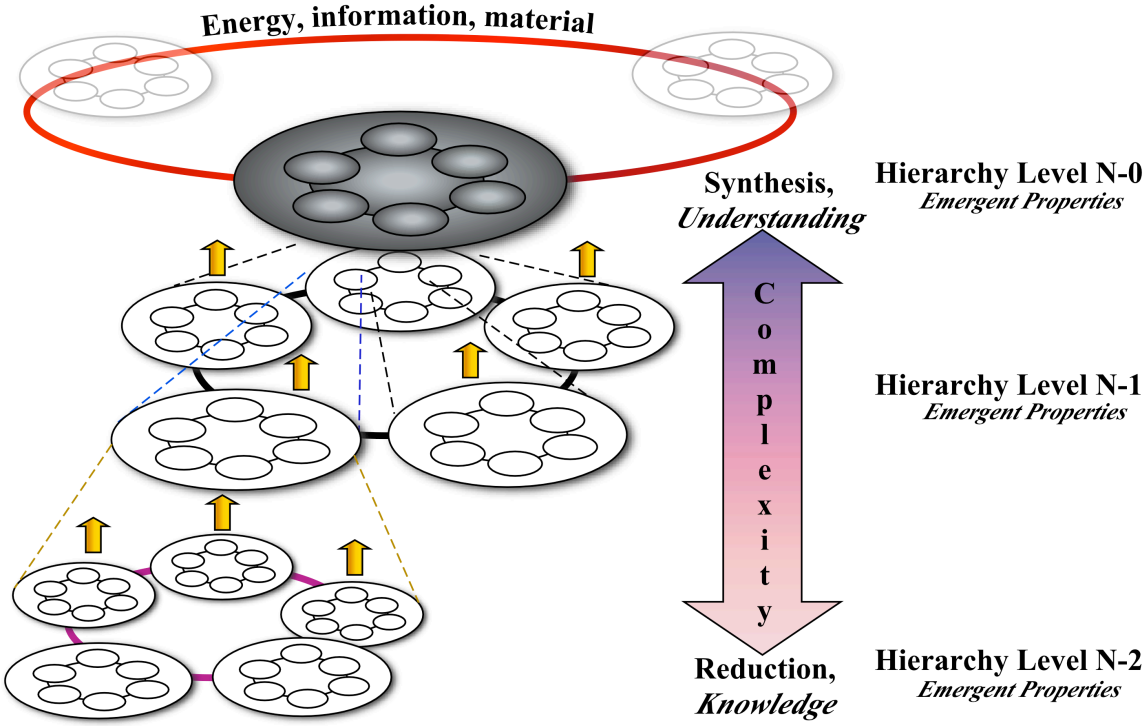


Figure 1. Growth of Complexity through Synthesis

The complexity arrow shows synthesis going up the hierarchy and reduction (analysis) going downwards. So, we can find out how things work and describe them by taking them apart, while we will understand their purpose, intent and behaviour by synthesizing/ observing the whole in situation and environment.

Evidently, integration of parts without regard for their mutual adaptation in situation and environment will prove inadequate – essentially, each layer in the hierarchy comprises the emergent properties of the lower layers, rather then their intrinsic properties. This suggests that simply adding body parts together, like Mary Shelley's Frankenstein, is unlikely to produce the desired result.

The Human Template

Happily, we know something of an exceedingly complex system that we might use – with care – as a template for any attempt to design an autonomous anthropomorphic machine.

The human body comprises some 11 or 12 organic subsystems, depending upon how you count them:

1. Integumentary – the outer layer of skin and appendages
2. Muscular
3. Skeletal
4. Nervous
5. Endocrine
6. Digestive
7. Respiratory
8. Cardiovascular
9. Urinary
10. Lymphatic and immune
11. Reproductive

Any autonomous machine is likely to find counterparts or equivalents to these organic subsystems. The integumentary organic subsystem has an obvious equivalent: the outer surface of the autonomous machine. Muscular and skeletal systems will also have self-evident equivalents. The nervous system of a human is exceedingly complex, including as it does the five senses, balance, and the brain, with its billions of neurons, which controls motor and sensor activities, but which also contains the higher centres of intelligence, judgment, etc. The endocrine system passes hormones into the bloodstream, controlling bodily functions and behaviour. While the nervous system controls fast response, the endocrine system operates on a slower and more lasting basis to maintain the stability of the body: the autonomous machine equivalent is less obvious, but may be associated with changed states of activity and operation, and with metabolic rate…

Items 6, 7 and 8, the digestive, respiratory and cardiovascular, are all concerned with energy – its supply, distribution, expenditure and disposal of metabolic waste. They will have an equivalent, presently undetermined and problematic, that may be quite different from their biological counterpart. Numbers 9 and 11, urinary and reproductive systems, are unlikely to require equivalents - yet. Number 10, however, one of the most complex organic subsystems outside of the brain, will surely require an equivalent to guard against, search out and neutralize pathogens, viruses and the like, and to protect against radiation and interference, to which any complex machine may fall prey.

The mutual interaction of all of these organic subsystems within their environment and situation somehow sustains life. Unlike most of our manmade/engineered systems, the organic subsystems are very closely coupled and intertwined, each depending upon the sum of the others for existence and performance. The organic subsystems interact both with each other and with their environment: for instance, personality (a human emergent property) develops partly as a result of interactions with other people, and partly as a result of interactions with other parts of the mind…and so affects, and is affected by, the health of the body.

Disappointingly, the list above tells us little about the whole human, its purpose, its capabilities, the functions it performs, etc. As Figure 1 indicated, analysis may provide knowledge of the parts, but it is synthesis that will give us understanding of the whole within its environment; the whole, as ever, is much greater than the sum of its parts.[3]

[3] Aristotle, 384-322BCE – Composition Laws

Looking at the whole human, it might not be unreasonable to consider the human body as a vehicle enabling our minds to carry out our purpose, to fulfil our ambitions, to pursue our missions, etc.

Already we can see that the synthesis of an anthropomorphic machine like an Autonomous Peace Officer (APO) would present two distinct challenges: the physical structure and the management of behaviour. The list of human organic subsystems above is of little use with the latter, which is the principal subject of this discourse.

The Autonomous Peace Officer [2]

The role of a human peace officer is well understood. S/he maintains and restores order in an area, often following a patrol route known as a 'beat;' this much has not changed since Talos, the ancient Greek Man of Bronze, supposedly patrolled Crete by circling it three times per day.

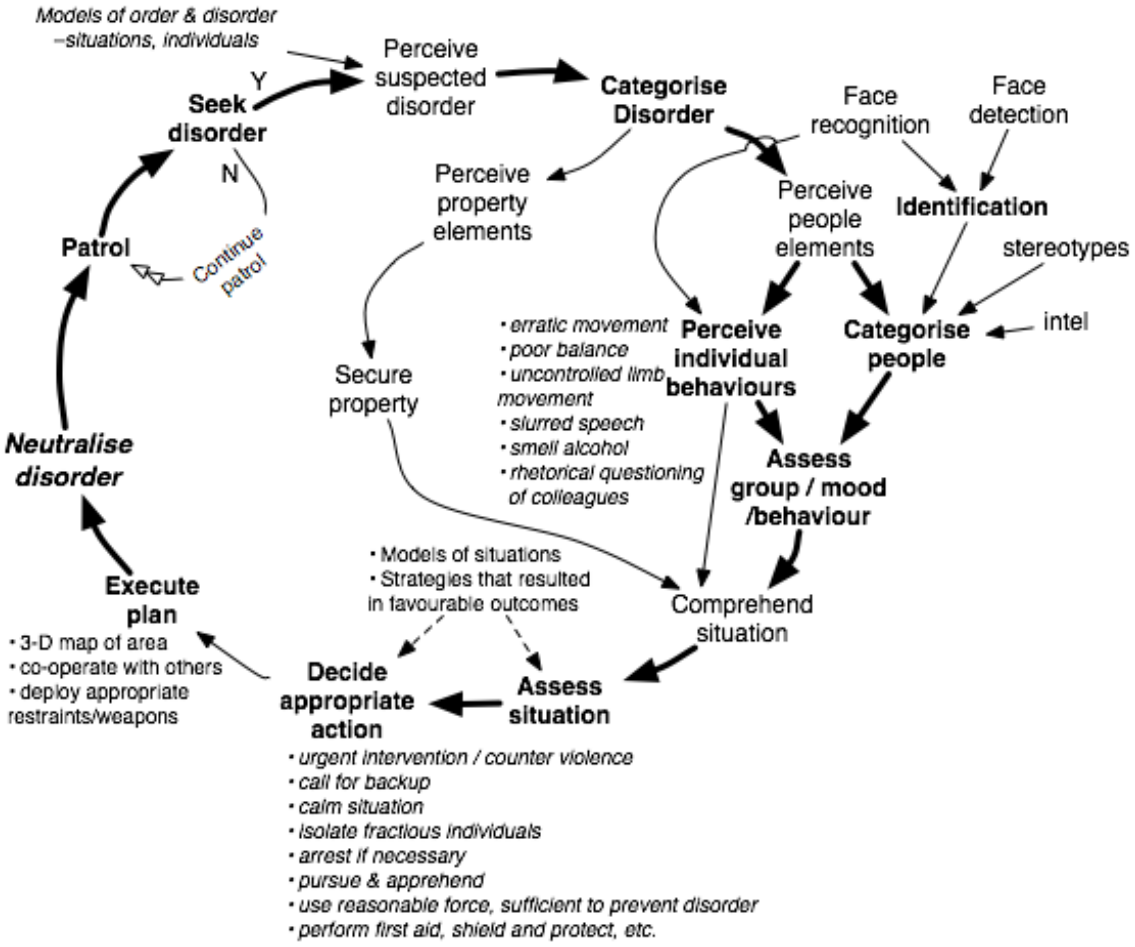


Figure 2. Typical Peace Officer Concept of Operations (CONOPS)

Figure 2 shows a typical Peace Officer Concept of Operations – there could be several others. To unfold the concept, start at Patrol – 10 o'clock on the continuous loop diagram. Following the diagram clockwise shows that the officer has to be able to categorize

disorder, categorize property and people, assess group mood and behaviour, assess situations, decide upon appropriate action, formulate and execute a plan of action to neutralize the disorder, after which s/he may resume patrol…

The peace officer is not alone in this work, although s/he may operate unaccompanied. Instead, s/he is a member of a policing team, with a command and control centre serving to coordinate the actions of the various peace officers, and if necessary to redirect officers in support of a colleague in need of assistance. Unlike soldiers, however, peace officers are expected to assess situations on their own and to make their own judgments. S/he is, after all, on the spot and therefore best aware of the situation…

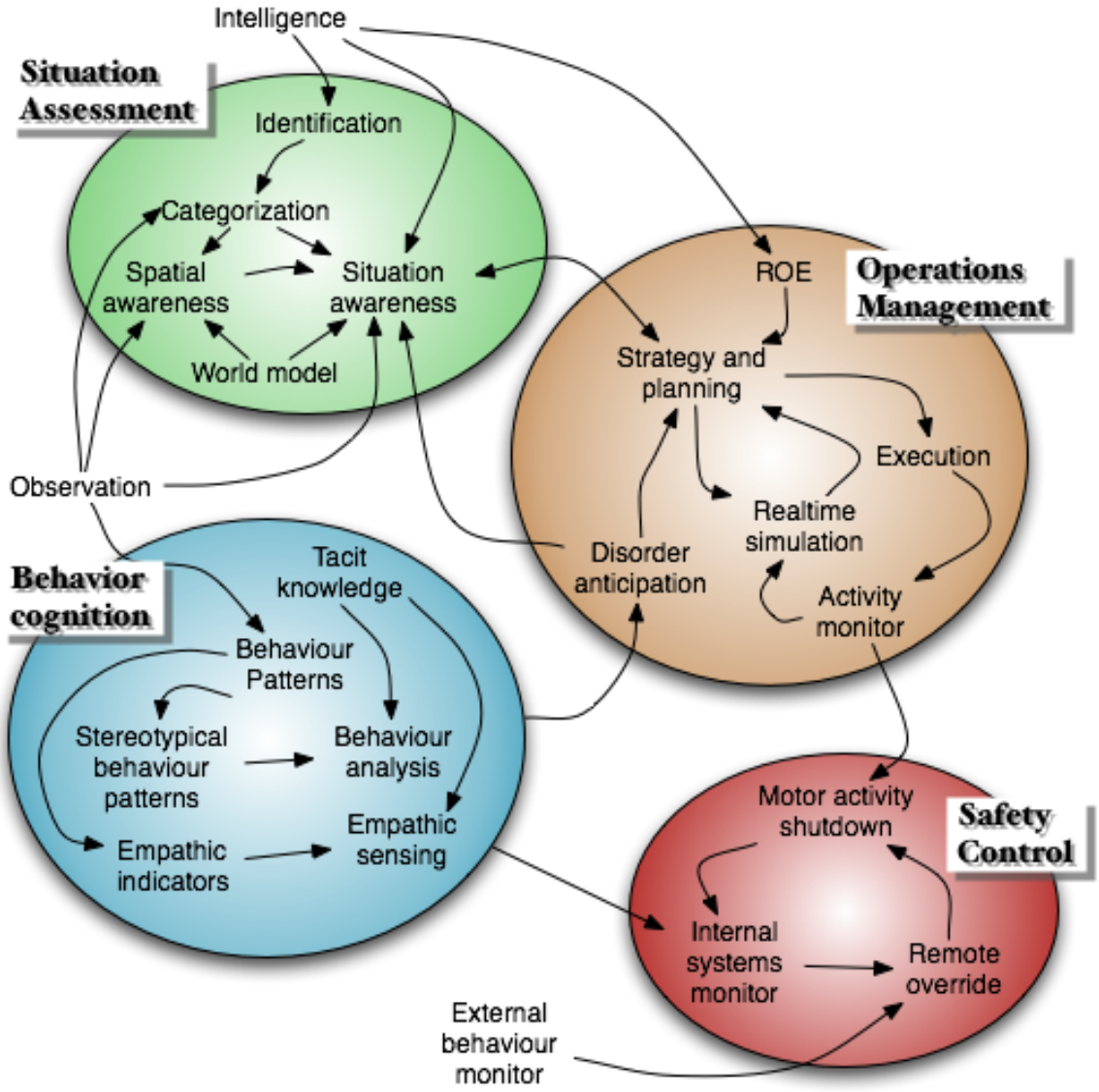


Figure 3. APO – Outline Cerebral Activities

Developing and operating to the CONOPS is largely cerebral, with aggressive physical activity arising only occasionally under 'decide appropriate action' and 'execute plan.' The question arises, nonetheless, how do peace officers go about assessing situations, for

instance, and is there any way in which an Autonomous Peace Officer might emulate them?

Figure 3 suggests several cerebral activities that an APO may be required to conduct:

1. *Situation Assessment*, where the APO brings together intelligence information, observational information ('people, places and things'), identification information, categorization information and world models (Weltanschauungen) sufficient to be 'aware' of a situation, i.e., know what is currently happening. 'Assessment' indicates degree and significance of disorder compared with some norm for the area, and suggests a prediction of likely change.
2. *Behaviour Cognition*, in which behaviour of observed individuals and groups is recognized and categorized. Peace Officers recognize 'uncontrolled limb movement,' for example, as a prelude to disorder and violence. Empathic sensing is included here: peace officers anticipate a person's situation, feelings (sic) and intentions from observing their behaviour in context compared with known contextual behaviours.
3. *Operations Management*, where Cognition of Situation leads to anticipation of disorder and the formulation of a plan of action – hence Strategy and Planning, within the framework of ROE – rules of engagement – such that the actions of the PO/APO are governed by politically sensitive, practical, ethical and legal rules. Within this area are conducted faster-than-real-time simulations of the likely outcome from any plan of action. Research shows that we humans do this all the time, even when crossing the road – we run a faster-than-real-time simulation to see if we can dodge the traffic, or not. We do it so quickly, and so subtly, however, that we are not consciously aware of it as a discrete activity – it blends into our pattern of behaviour. So must it be with the APO, where speed will similarly be 'of the essence…'
4. Finally, *Safety Control*. Items 1 to 3 may be rational, logical and sensible, but there is always, somewhere, the concern that such behaviour will result in an irrational or dangerous outcome, so some form of real time 'conscience' is required. In the APO, this might take the form of an overriding, and quite independent, way of assessing the outcome of the deliberations, and effectively determining: "the outcome may be rational and logical, but it is nonetheless, inhumane/politically insensitive," and so shutting it down.

Going into a little more detail brings into clearer focus the kinds of cerebral activities that an APO will need to undertake. Figure 4 outlines continually repeating processes that an APO might execute, conceptually. The box at the left of Figure 4 shows Situation Awareness as an outline, high-level process, leading to Recognition-Primed Decision Making in the box at right.

Situation Awareness shows the use of face detection techniques, such as we find in digital cameras we can purchase on the high street. Face and smile detection are now commonplace in small cameras, and facial recognition, while still in its infancy along with body and limb dynamics, is developing rapidly – it depends, of course, on there being a database of faces amongst which to find, and correctly recognize, the appropriate face, and so to identify the person. Happily, people who need to be identified are more likely to be in the database… this is much the same as the human

Peace Officer visually recognizing a villain from the 'mug-shots' in the station, and anticipating trouble accordingly.

Interactions with the public take place in an environment with which the APO would be familiar through training; he would be equipped with a 3-D map of the area and Satnav-like sensors to locate him in position and orientation, so that he may optically map his visual perceptions (people, places and things) and cues on to his 3-D map.

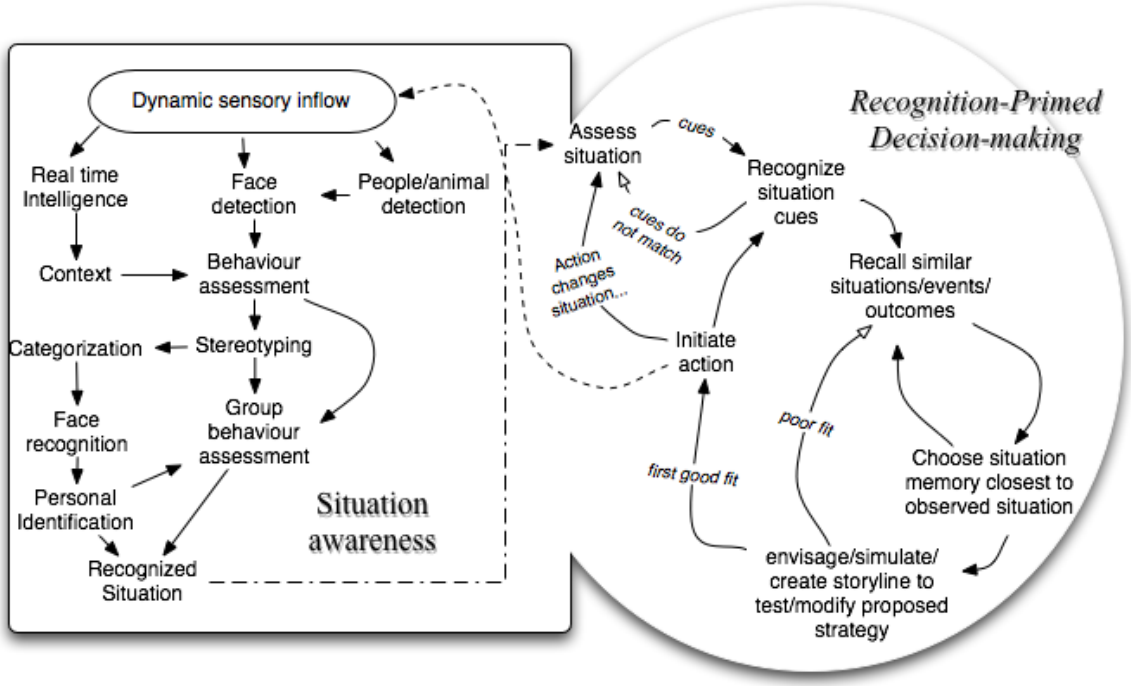


Figure 4. APO: Strategies, Planning and Decision-Making

Recognition-Primed Decision Making (RPD) [3] is the very fast decision-making made by experts under time pressure. While the APO does not fit that bill, he may be provided with a database of previous situations and corresponding strategies that proved successful: he may build upon that information as he becomes more experienced in the field, i.e., he may update his database. Meanwhile, upon assessing a situation, he will be able to 'recall' similar situations, events and their outcomes, and hence derive a strategy for dealing that is most likely to succeed.[4] Note too, that Figure 4 is a continuous activity, not a one-off, so assessments, decisions and strategies may be continually updated as the situation unfolds: the expert decision-maker does not make an initial decision and stick with it; instead s/he will update that decision as and when new data becomes available, will re-assess, and so will progressively 'home-in' on the best choice.

APO Behaviour Management

Figure 5 shows a general model of how behaviour may be managed [4] conceptually in humans, after Karl Jung [5]. It incorporates Nature Vs. Nurture, in the grey panels, and invokes a Belief System. The arrow and list, at top right of the figure, influence evolved Nature; while Belief System incorporates ideas indicated by the arrow and list, at lower

[4] The data rates and volumes needed to emulate an expert decision-maker are likely to prove challenging.

right. The model follows a classic stimulus response paradigm. Stimulus enters centre left, into Cognition. A stimulus is 'recognized' or interpreted by reference to tacit knowledge, world models and belief, so that an *interpretation* of the stimulus emerges which may, or may not, be accurate, according to familiarity with the stimulus, and what was expected in the current situation (belief).

The interpretation of stimulus is passed to (behaviour) Selection, where Nature, Experience, Belief and constraints (such as capability, opportunity) all influence the selection of Behaviour, from a range of 'primitive' behaviours (aggressive, defensive, cooperative, cautious, etc.), in response to stimulus. Nature, too, presents long established archetypes[5]: king, magus, shepherd, knight, etc., which determine the way in which individual behaviour 'emerges.' Often in humans, nature dictates instant response for survival, ("knee-jerk"), which may be swiftly overridden by considered judgment. Note, too, the importance of training in this respect, indicating that both people and suitably designed APOs may be trained to behave/respond in ways that are contrary to their natures.

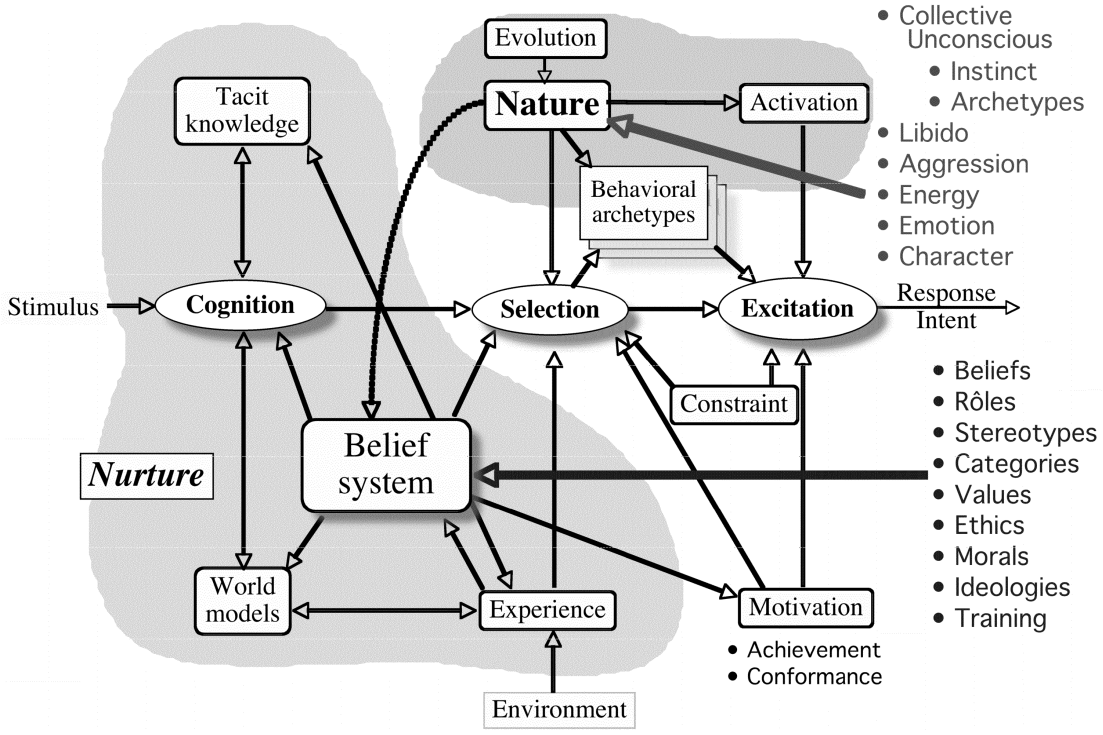


Figure 5 Management of Behaviour

The selected archetype-influenced Behaviour is then passed into the rest of the information-decision-action procedure, where it influences the selection of options, choice of strategy and control of implementation that are the normal stuff of command and control.

For the APO, there are serious challenges in establishing Tacit Knowledge [6] (simple everyday facts like grass is green, sky is blue, things fall, etc.), World Models and Belief

[5] For an PO/APO, a suitable combination of Jung's archetypes might be "Shepherd-Knight," reflecting his dual role as protector of his flock, and knight challenging would-be aggressors, according to situation.

Systems. Vast amounts of such information would be needed to create a general-purpose autonomous machine (e.g. Data on Star Trek.)

However, it may be practicable to create appropriate sets of tacit knowledge and world models for an APO in a predetermined, confined situation on a given beat. Similarly, any general belief system would be exceedingly large and complex, but the belief system associated with an APO in a particular location/situation/beat may be feasible; for instance, the number and variety of stereotypes may be quite limited. (For policing, stereotypes are invaluable – if you dress like a vagabond, chances are that you are a vagabond, and you should not be too surprised to be treated as one – unless, that is, there is a local carnival in progress – context changes everything!).

Such is the complexity likely to be involved with each of tacit knowledge, world models and belief systems, that it may prove more practicable for the APO to develop these over a period of time, some perhaps in simulation, some in real situations, and some pre-programmed. Similarly, the APO may have to learn how to stand, balance, run, jump, fall

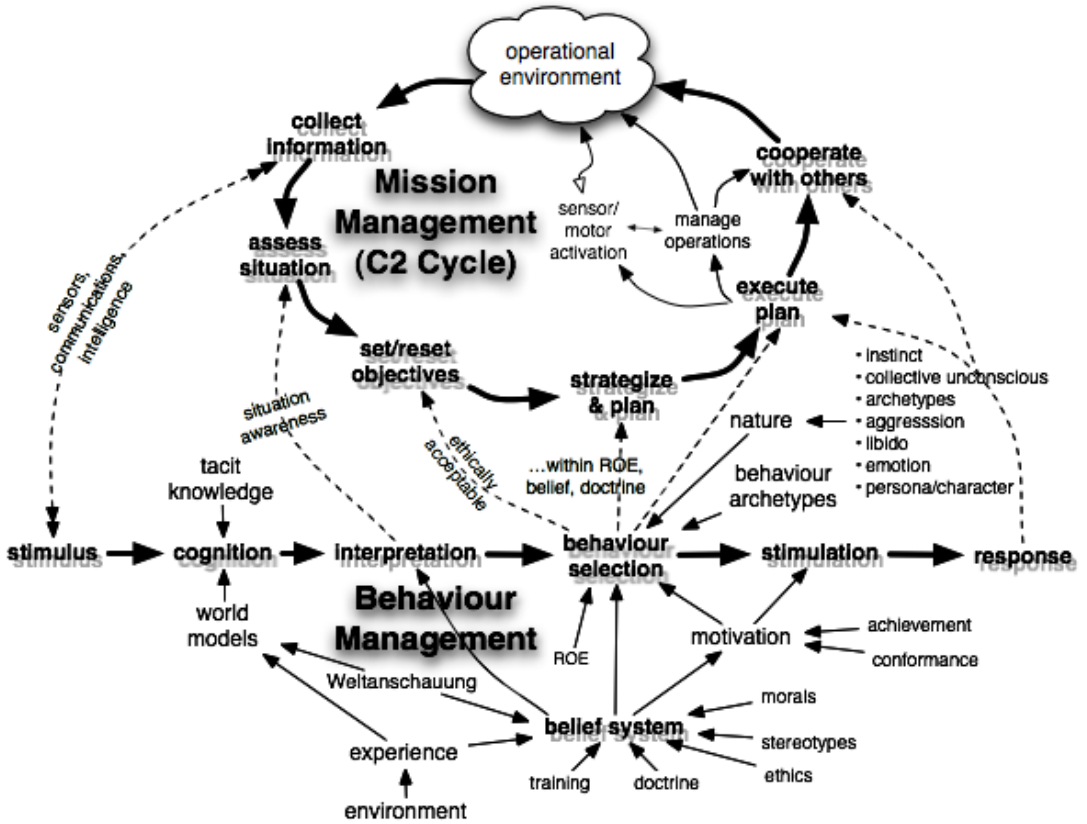


Figure 6. A design 'framework' on which to 'hang' the APO 'cerebral' design and architecture. The upper 'circle' forms the well known command and control (C2) cycle widely employed in mission management. The lower section goes from stimulus (left) to response (right), showing how behaviour is invoked as response to stimulus. The two sections are connected, indicating ways in which behaviour influences the C2 cycle. The framework shows logical functions, datasets and relationships, but need not be configured in the best way for particular functional architectures, e.g. for situation awareness/assessment. The whole is a complex synthesis of systems…

over, recover, etc… in order to make these activities efficient and life-like. Creating such abilities in automatons is currently proving challenging…

Autonomous Peace Officer Functional Design Concept

Figure 6 shows the behaviour management model of Figure 5 coupled with a more conventional command and control (C2) loop at the top typically used in mission management: collect information; assess situation; set/rest objectives; strategize and plan; execute plan; and cooperate with others if necessary in the process. Information is collected from the operational environment, and executing a plan is likely to change that operational environment, so this is a continuous loop, responding to, and initiating change. Moreover, it may be addressing a number of concurrent missions, so loop activity can be frenetic…

On its own, the C2 loop is designed to produce rational decisions in dynamic situations, which may in some circumstances be considered 'uncompromising' – a charge occasionally levelled at the military. However, in the APO Design Concept, the C2 loop is moderated by Behaviour Management, so that training, ethics, morals, beliefs, Rules of Engagement, etc., can be brought to bear on the decision-making process. Note, too, how the belief system becomes involved in situation awareness, how objectives become ethically acceptable, how strategies and plans operate within doctrine, ROE, and Jung's Archetypes, e.g. Shepherd-Knight for an APO. In some situations, there are decisions that would fall naturally to the *shepherd of his flock,* which must be protected, or to the knight, *champion of his people*, who must behave chivalrously.

At top right there is a small, but important, loop: *manage operations*. We humans are able to continue assessing situations and making judgments even while pursuing some objective. It seems that the pursuit of the objective is conducted subliminally while our minds are looking ahead, anticipating problems, developing new tactics, etc. This is an important feature for an APO too: that he be able to conduct current operations while assessing developing situations. The loop *manage operations* indicates that the design delegates current operations to a lower level processor, so that higher-level situation assessment and decision-making may continue uninterrupted and at speed.

A major question remains unresolved: how can the 'cerebral' processes indicated in Figures 2 'engage' with the physical elements of the complete APO to make an operating machine that can run, jump, perceive an offender, give chase, restrain if necessary, interrogate, and – most importantly for any Peace Officer – apply *discretion*. Discretion marks the Peace Officer as "reasonable" – the peace officer is able to use his judgment about which minor misdemeanours he prosecutes and which it is reasonable, in the circumstances, to let go. So, must it be for the APO – he must be able to apply discretion, else be dubbed a 'dumb, insensitive machine'…

The outline problem is illustrated in Figure 7: sensor, motor and execution coordination feed into the primary mission functions, supporting purpose. But, how do the cerebral activities of mission management and behaviour management actually control the physical elements of sensor, motor and execution: there is a lacuna – the grey loop in the figure represents the uncertainty of connection?

We can get some idea of how to connect cerebral purpose to physical realization by considering a simple example. Prime Mission Functions for the Peace Officer are indicated in Figure 2. Consider 'Pursue and Apprehend.' The APO may be seated when he (we will anthropomorphize the APO as male) observes a miscreant running away and a person shouting for help: the APO elects to apprehend the miscreant, so establishing a new mission. He will then plan an intercept path, avoiding people and objects, as he stands and gains balance – both tricky operations, requiring a considerable amount of coordinated muscular activity in legs, back, torso, arms and head: simply standing from a low seat requires humans to rock back and the forward, thrusting the head and arms forward while pushing with the legs, straightening the back, tightening the stomach muscles, and moving body weight on to the toes – it is surprisingly complicated to stand and at the same time move forward smoothly into running mode without losing balance, yet we do it without conscious thought...
[For the APO each of these elemental functions will be learned routines, each routine triggering complex sequences of 'muscular' actions: a routine for standing in balance; one for transition to running; one for running in balance, etc., etc. just as in a human, and similarly subliminal…]

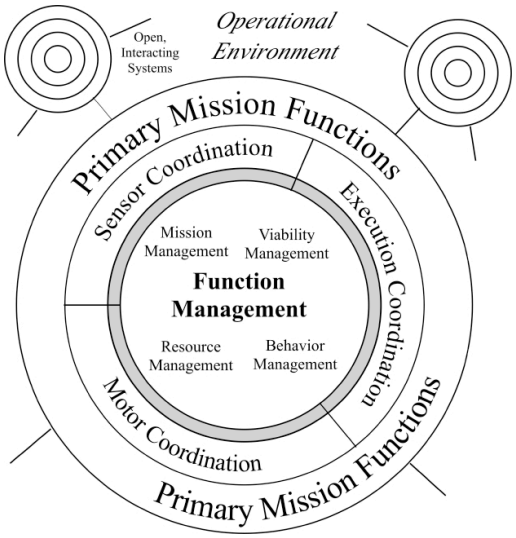


Figure 7. Connecting Purpose to Action

The APO, having achieved this feat, will then run on an intercept path towards the miscreant, who may dodge, speed, up, slow down, requiring the APO to continually re-plan his intercept path, swerve, jump… Assuming that the APO is nimble and fast enough to catch up with the would-be escapee, the APO has then to restrain the miscreant without harming in her or him. This is a difficult manoeuvre, and may require the APO to deploy some non-lethal weapon, such as a "sticky net," to avoid cries of 'police brutality.' (If a human were killed by an APO, intentionally or inadvertently, the political fallout would be tremendous.)

Having overcome the restraining hurdle, the APO has then to subdue the miscreant, question him/her, seek to identify them, report into command central, recover any weapons or stolen goods, and so on. All of these would be serious challenges to an APO

in the real world, and so far we have considered only one of the Prime Mission Functions from Figure 2. Note, however, that each PMF is carefully and extensively coded for human Peace Officers, which limits the range of actions required of the APO.

Figure 8 shows how various routines and sequences might be organized. Starting at top left of the diagram is the formulation of a Plan to execute some Primary Mission Function. Following the arrows leads to an Action Semantic Map that shows various sensor and motor actions being activated to constitute a routine. The output from the Map leads to a second map that transitions from sequence to sequence of routines, some in series some in parallel, the whole progressively executing the Primary Mission Function. The whole appears – and is – potentially complex, since it has to be able to transition smoothly from routine to routine, for example, as the miscreant swerves and dodges…

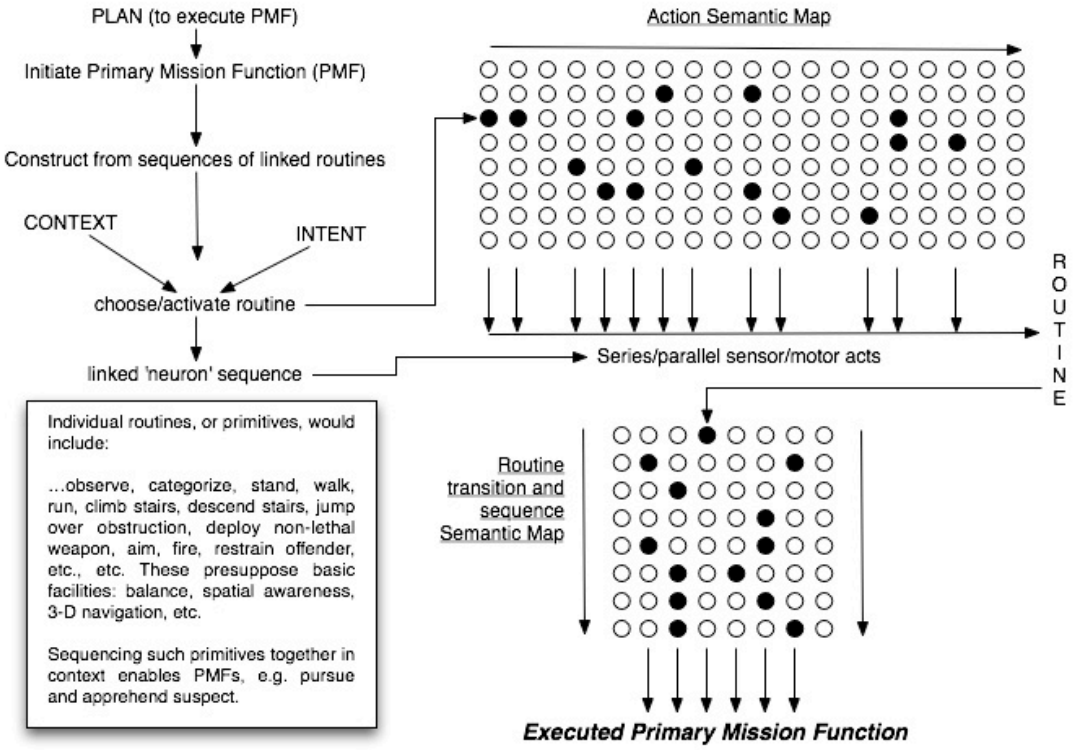


Figure 8. Synthesizing a Primary Mission Function

The APO may also find himself required to sympathize with a victim of some crime, so that different modes of speech and behaviour will be needed, according to situation, in accord with the appropriate behavioural archetype (e.g. Shepherd-Knight). Some fascinating work with chimpanzees by Rizzolati et al [7], suggest that it may be possible for the APO to empathize both with victims and would-be perpetrators, by comparing their behaviour in context with a database of known/observed contextual behaviours, and so to anticipate likely outcomes. The APO's ability to exercise discretion is likely to lie in his ability to empathize with miscreants: to assess from their attitude from behaviour-in-context (anger, aggression, escape, regret, guilt, remorse); and, simultaneously knowing their misdemeanour records, to make a judgment call about prosecution.

An APO design feasibility study [2] indicates, then, that there is a phenomenal amount of data to be amassed to represent tacit knowledge, world models, and belief systems. Furthermore, an APO should be able to access, review, select and employ appropriate data very quickly to be practical for interaction with people – c.20ms is a sensible target, one easily reached by humans, and an APO could not afford to be slower. In developing strategies and plans, too, there would be a need to conduct simulations of proposed activities in context, to prove their effectiveness, and to mitigate risk to life and limb. (Estimates suggest that processing speeds in the order of 10^{11} bits/sec might be needed to achieve the necessary response times for Recognition-Primed Decision Making alone…)

In many situations, the APO would be called upon to discuss situations with both victims and suspects, and to make reasoned and 'instinctive' judgments as to the merits of a case: this would require more than a facility with language; as well as being fluent, the APO would need to be able to converse purposefully, to formulate sensible questions in context, and judge the validity of responses.

The Systems Design Concept Outline

Figure 9 gives an outline of the more straightforward aspects of design, excluding such

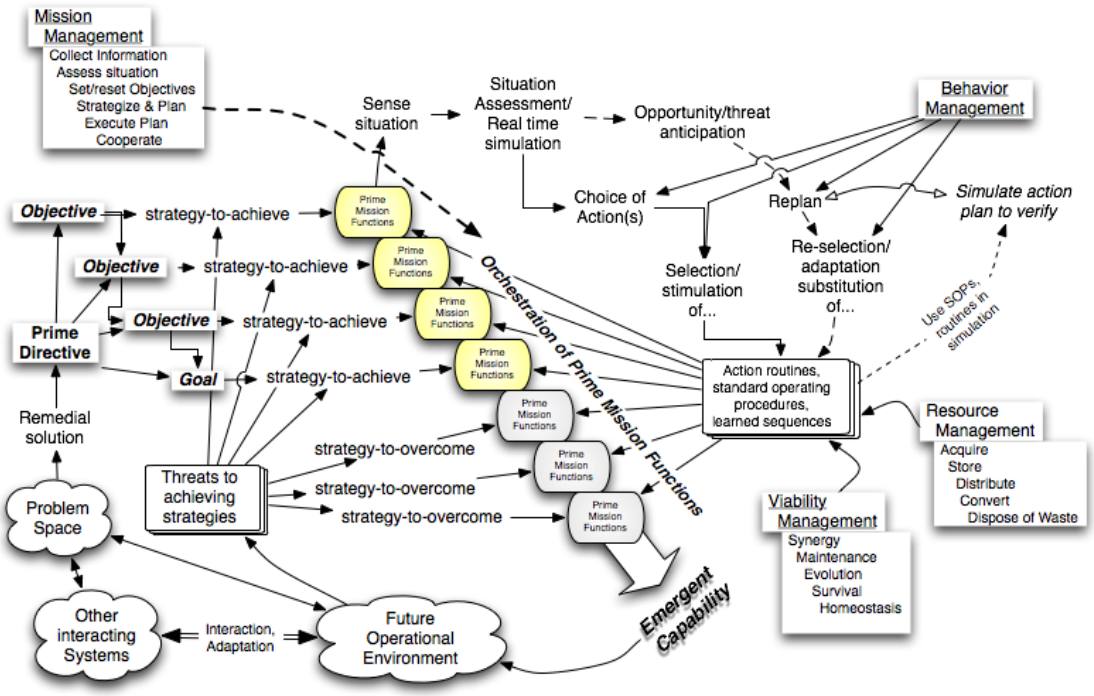


Figure 9. The System Design Concept.
The objective is to create 'Primary Mission Functions," that delineate the APO's capabilities, and to develop these by identifying and orchestrating a set of action routines, operating procedures and learned routines. The APO needs the ability to sense situations (top), assess them in real time, choose appropriate actions, simulate them to test their potential, and stimulate them in sequences that create each PMF. As the figure shows, there is a circular process resulting in closure, as the activated PMFs create emergent properties, capabilities and behaviours in the Future Operational Environment (bottom) which resolve the various problems and issues within the Problem Space. Each PMF 'emerges' from the synthesis of many subsystems interacting in cooperation and coordination, forming a complex, ever-shifting pattern of activity.

problematic higher-level APO cerebral activities. The line of Prime Mission Functions runs from top left of the figure to bottom centre. The Prime Directive (shown at centre left of the figure), is the ultimate statement of purpose, which – for an APO – might be "to maintain public order." The Prime Directive suggests a Goal (societal order?), which can be achieved through a series of objectives (achieve: property order; individual order; group order; societal order – hence the Goal).

Achievement of each objective and of the Goal necessitates the generation of an appropriate strategy, and there will inevitably be threats to achieving some or all of those strategies. Together objectives, goal and strategies indicate the range of Prime Mission Functions that the Solution System (APO) will perform. As the figure shows, the execution of these Prime Mission Functions in the Future Operational Environment will present an Emergent Capability, constituting the Remedial Solution…

Emerging from the top Prime Mission Function, observe Sense Situation, leading to Situation Assessment/Real-time Simulation, and moving into the command and control activities indicated by the Mission Management panel, top left. Behaviour Management, top right, influences choices and decisions as Figures 5 and 6 indicated, leading to an aggregation of action routines, standard operating procedures and learned sequences that together enable the autonomous system to 'animate' the Prime Mission Functions appropriate to the immediate situation.

Note that the figure represents a continuous process, a control loop, where the characteristics of the Problem in the evolving Operational Environment automatically generate and apply the necessary solution – hopefully in real time.

Note too the Viability Management and Resource Management Panels, (part of the Generic Reference Model, which purports to show the 'inside' of any complex system [8]) which contribute respectively to the continuing fitness of the Solution System and to its being continually resourced…including the vital topic of energy. Viability Management, in particular, addresses much of the physical side of the solution system: synergy, maintenance, evolution, survivability and homeostasis. In some autonomous systems it may be necessary for them to *evolve* by sensing/improving their performance in changing environments, to be *self-healing,* and to be *survivable* (which will include damage repair). *Homeostasis* for such a self-evolving, self-healing complex is likely to be an issue in a dynamic and constantly evolving societal situation. Their behavioural archetypes (Shepherd-Knight, etc) would remain unchanged, however, serving as a conscience and controller of excess behaviour…

APO conclusions

Organizations around the world have addressed different aspects of androids which, brought together, might conceivably provide the basis for a working APO, and some surprising developments such as face detection have emerged, which start to change the environment. While it might be possible, in theory and in principle, to bring all these separate advances together as proposed in the APO design concept, the resulting synthesis would be indeed complex and, since the machine would inevitably learn and adapt to situation and context, quite how it would behave would be difficult to predict in advance. Perhaps it would be excellent…perhaps limited.

In the real world, the police officer is frequently called upon to deal with events and situations that are not represented in Figure 2: a major fire; an airplane crash; a drug bust; an earthquake; etc. Is it feasible to equip an APO with the ability to deal with the unexpected? But then, the human peace officer would have his/her limits too. There is great value in conducting feasibility design studies of the APO, if only to remind us just how special humans can be…and the limitations of technology by comparison.

Finally, would it be *wise* to build and deploy an APO? There is no evidence that a 'humane' APO as described above would be acceptable, although philosophers seem to think so – Isaac Asimov's *I Robot* being a case in point. He, the APO, may be accepted, might even be considered preferable to human peace officer by virtue of impartiality, ethical behaviour, discretionary adherence to the law, survivability in harsh environments, etc. On the other hand, he may be considered weak, or slow, or stupid, and so ridiculed which may challenge his authority as an officer of the law. He might become the butt of practical jokes, testing for his weak points vis-à-vis a human peace officer, and looking on his 'discretion' as weakness after the style of teenagers.

Overall then, perhaps not wise. But the APO will happen. Soon. Like Pandora's Box, it is irresistible…

Autonomous Air Vehicles (AAVs)

Different Domains, Different Objectives

Autonomous Vehicles are a contemporary reality for military applications, where risks and constraints are different from civilian and commercial life, in which people-concerns generally preclude full autonomy. The Jubilee Line on the London Underground is capable of unmanned operation, but operates manned owing to safety concerns by the travelling public and unions. Similarly, air transport could, in principle, operate unmanned in many cases, but the fare-paying public are much more comfortable with a real, fallible person at the controls than a supposedly infallible machine…

Autonomous Ground Attack

Autonomous air vehicles are used in Afghanistan for attacking ground targets. AAVs do not risk the lives of the attacking force, a very significant advantage to the attacker. However, there are the ever-present risks of unintended collateral damage and civilian deaths. The latter, in particular, can prove counter-productive, stirring up resentment and hatred, and firing-up a new army of jihadists, quite the opposite of the intentions.

So, would it be practicable to equip such AAVs with the ability to:

- Assess the situation upon arrival in the target area
- Acquire real-time intelligence about the intended target
- Assess the potential for civilian casualties and collateral damage
- Determine, as could a human pilot, whether it was prudent to proceed in the circumstances
- If not, to search around for another target (an 'alternate' may have already been designated) and to assess its viability as a reasonable target
- Etc.?

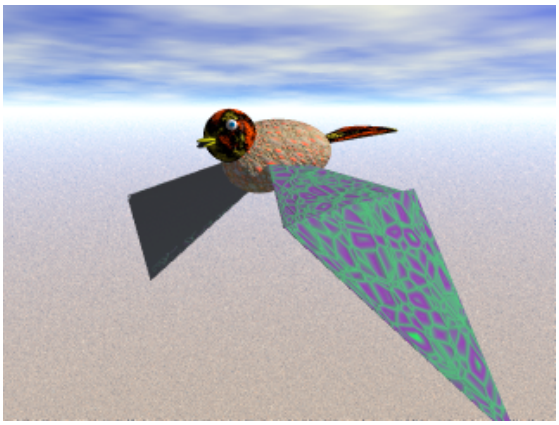
Even if it were practicable – and it may be - there is an immediate problem. If an AAV were equipped with the 'behaviour,' to make humane judgment calls, and the opposition came to know of it, then the opposition could invoke 'human shields:' civilians, hospitals, schools, etc. This tactic, towards which several countries have already shown a propensity, negates the attackers' intent to be humane.

The attacked regime essentially dares the attacker to attack such sensitive targets, and will assuredly use any such attack – whether against a real, or mock, target – as a publicity weapon. It is, assuredly, an effective counter. The attacker is left with the need to improve their intelligence, introduce more selective weapons, or just call the bluff and attack the target willy-nilly and live with the political fallout.

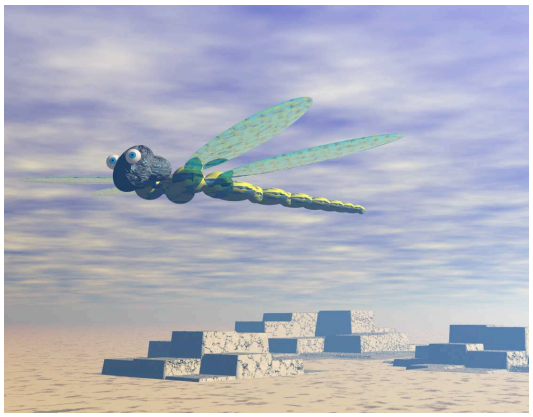
An Alternative Approach

In some situations, however, there may be a different way – conceptually, at least, and depending upon situation, context and – in this instance – technology. Consider a potential combat situation [9] where an invader was entering an area that was ecologically sensitive, where damage to the environment could be immense. There are many such areas on Earth at present: deserts, Great Plains, Polar Regions, etc.

Consider further that a transportable mobile force, Land Force 2020, had been established to defend the territory, to eject the invaders, but with minimal damage to sensitive ecosystems and with little loss of human life, on the part of both defenders and invaders. Such a force might arise under the auspices of the United Nations, for example.



The raptor, presently uncovered, will be 'dressed as an indigenous bird of prey, with flapping wings, capable of finding and soaring on thermals: wings are solar panels, tail is an antenna, and eyes are video cameras. Raptors are armed and carry dragonflies.



The dragonfly, much smaller than the raptor, similarly has solar panel wings and video camera eyes. The body is an antenna. Dragonflies can operate individually or in swarms, can hover, dart and dodge, and are armed.

Figure 10. Land Force 2020 Raptor and Dragonfly.

The force is comprised of lightweight land vehicles, transported by global transport aircraft, so that they may be deployed swiftly, but need reside in territory. The land vehicles are not designed to fight, but are able to move across all kinds of terrain and water, operating in automatically controlled formations, swarms, according to situation and context. There may be one human operator to several automatically steered vehicles, greatly reducing the risk to human life.

Each vehicle carries a number of autonomous solar-powered *raptors*, which may be launched and retrieved on the move. Each raptor carries a number of *dragonflies*. The raptors are able to soar on thermals, so may stay aloft for hours or days, using very little energy. They are made to look like indigenous birds of prey to avoid disturbing the local fauna and alerting the invaders. Similarly, the dragonflies are made to look like, uh, dragonflies…Figure 10 shows dynamic simulation models of raptor and dragonfly.

Envisage a mobile force, a swarm, of land elements travelling across open country, preceded by their autonomous raptors, with the raptors launching autonomous dragonflies as needed. Raptors and dragonflies are equipped with video-camera eyes and digital communications, allowing any human controllers to detect, locate, identify, monitor and even communicate with any intruders. There could be a command and control centre with several operators in the supporting transport aircraft, which would stay in the rear.

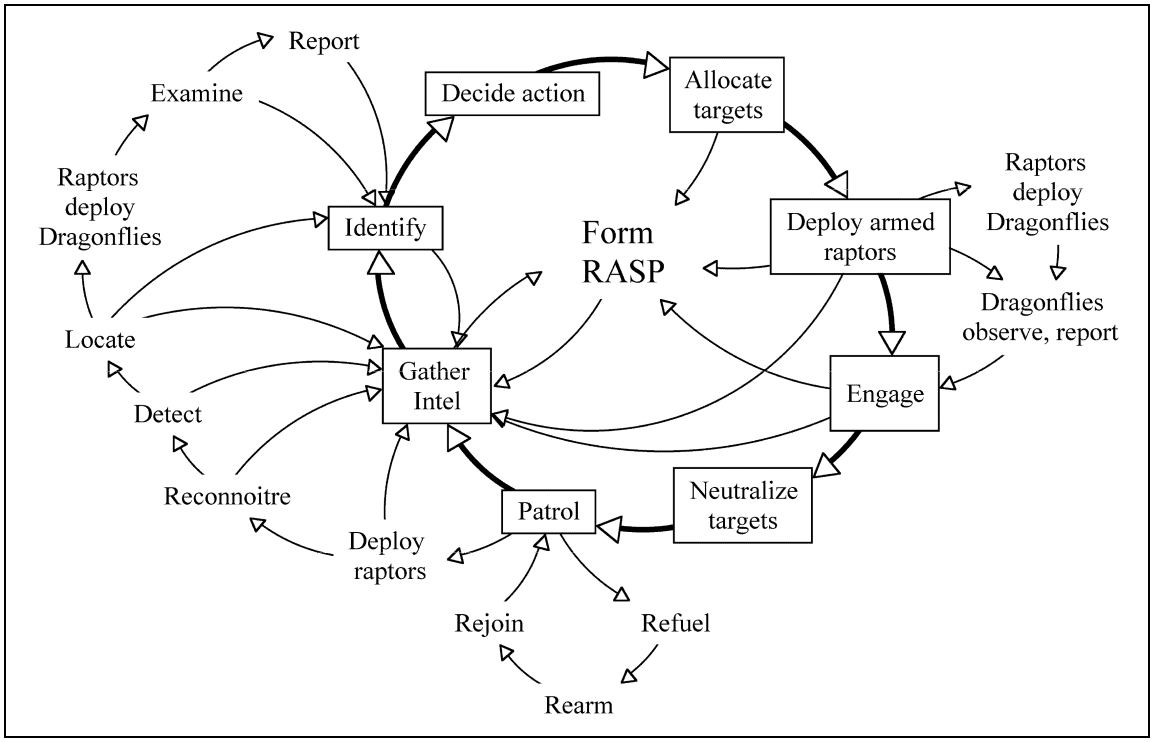


Figure 11. Land Force 2020 CONOPS

Figure 11 shows a conceptual CONOPS for such a situation. Note the resemblance between the main loop in the figure and that of Figure 2 above, the Peace Officer CONOPS. As Figure 10 shows, the swarm of land vehicles travels with raptors ahead of

them reconnoitring, and following the routine: detect; locate; identify. Intelligence may be passed back to C2, but – if straightforward – may be addressed by a raptor, which may deploy a dragonfly to investigate. A RASP (Recognized Air and Surface Picture) is formed using data from all sources, and the RASP is available to C2, Land vehicles and raptors alike. The RASP integrates and presents the surveillance, acquisition, targeting, and kill/deter assessment (SATKA) of the whole force.

A decision to act on Intel may be made by the land vehicle commander, the C2 in the supporting aircraft, or automatically (in the absence of command) by the raptors together – these would be network-centric operations. Action may include raptor and dragonfly weapons, but the dragonflies are able to address and converse with intruders – acting as '2-way-comms' links with C2, so intruders may be warned, may be advised to stand down or withdraw. If conflict in inevitable, then, raptors and dragonflies are equipped with a variety of lethal and non-lethal weapons to incapacitate the intruders, and/or render their technology useless. Dragonflies are potentially able to get very close to, hover, pick out a particular human target and to neutralize that target without risk, let or hindrance to other people, places or things…

Is the concept viable?

Surprisingly, perhaps, the concept is viable in the near-to-mid term, drawing upon current research in a number of areas for autonomous control of land vehicles, for the creation of fast-beat mechanical wings for small, agile, unmanned air vehicles, and for non-lethal weapons. Communications are based on mobile smart phone technology, which is currently able to communicate visually and aurally and could, with some modification, form the binocular eyes of both raptor and dragonfly. Bringing all of these together and creating the whole solution would be expensive, and at present no one appears to have either the will or the money.

However, the systems design concept does suggest that the behaviour of our autonomous fighting systems could be made more effective, subtler, less aggressive, and less harmful to sensitive environments in future – if we had the will… In effect, the Land Force 2010 concept is of a 'natural' extended system, with a distinct 'whole-system' behaviour…

Consciousness and Sentience

Note the absence of any reference to machine consciousness, or sentience. [10] Whether an APO that was aware of, and able to assess situations, identify threats to order and make fast, Recognition-Primed Decisions to restore order could be deemed sentient or conscious, may be unfathomable. Perhaps there are degrees of consciousness. Are we humans fully conscious at all time – whatever 'fully conscious means' – or are we more conscious at times of high alert, less conscious at times of relaxation? The APO would not recognize the questions, let alone the implications…he would be unable to think about himself as an entity, although he would be aware of his own capabilities and limitations within his narrowly defined domain context… within those constraints, might he be deemed 'self aware?'

Conclusion

Autonomous machines of the future are widely feared for their anticipated inhumane behaviour: literature and the media have stoked fear, to such an extent that the general public may discount the potential advantages of autonomous systems in support of humanity.

The idea is gaining some ground is that we might endow our autonomous machines with more human characteristics – appropriate facial expressions, personality, empathy, appropriate human-like behaviours, morality and ethics – so that we humans will find machine human interactions less daunting, more rewarding, while the autonomous machines will be less likely to 'run amok' or behave in a cold, impersonal, machine-like manner.

It seems reasonable to suppose that we could so endow autonomous machines, at a cost: yet, there is no hard evidence that autonomous machines so endowed would really be more acceptable – indeed, it is difficult to see how such evidence could be garnered without practical examples and trials. However, it seems *reasonable*…

Exploring the practicality of endowing autonomous machines with human-like behaviour suggests that current technology may not be sufficient to do this in the general case, but that it might be possible, by constraining the domain of use and operation of the autonomous machine, to bring the technology required within attainable limits.

The likely cost of doing this also suggests that trained humans are likely to remain the best (most cost-effective) solution in cases where danger and extremes of environment are not the driving considerations…

Meanwhile, innovative technological advances such as face detection, smile detection, facial recognition and improved understanding of empathy on the cerebral side, and the development of small, agile flying machines on the physical side, together with the potential repurposing of current technology such as smart phones, suggest that there may be alternative ways of creating smart, affordable, autonomous machines that we have yet to think of… So, perhaps, the notion that 'humane' autonomous machines are too complex and expensive is being eroded by the ever-rising tide of technological advance.

Derek Hitchins June 2011

References

[1] Hitchins, D K, (2000) Getting to Grips with Complexity…A Theory of Everything Else, www.hitchins.net/ASE_Book.html

[2] Hitchins, D. K, 2007 IET Autonomous Systems Conference, *Design Feasibility Study for an Autonomous Peace Officer,* http://www.hitchins.net/Systems%20Design%20/AutonomousPeaceOfficer.pdf

[3] Klein, G.A. (1989) *Recognition-Primed Decisions,* Advances in Man-Machine Research, 5, JAI Press

[4] Hitchins, D.K., (2003), *Advanced Systems Thinking, Engineering and Management,* Artech House, Boston MA

[5] Jung, C. G., (1976), *The Portable Jung*, Penguin Books.

[6] Polanyi, Michael.(1983) "The Tacit Dimension". First published Doubleday & Co, 1966. Reprinted Peter Smith, Gloucester, Mass, Chapter 1: "Tacit Knowing".

[7] Rizzolati, G., Fogassi, L., Gallese, V. '*Mirrors in the Mind,'* Scientific American Vol. 295 No.5, Nov. 2006

[8] Hitchins, D.K., (2007), *Systems Engineering: A 21st Century Systems Methodology,* pp.124-142, John Wiley & Sons, Chichester UK

[9] Hitchins, D.K., (2007), *Systems Engineering: A 21st Century Systems Methodology,* pp.313-348, John Wiley & Sons, Chichester UK

[10] Koch, Christof and Tononi, Giulio, A Test for Consciousness, Scientific American, Vol 304, No. 6, June 2011