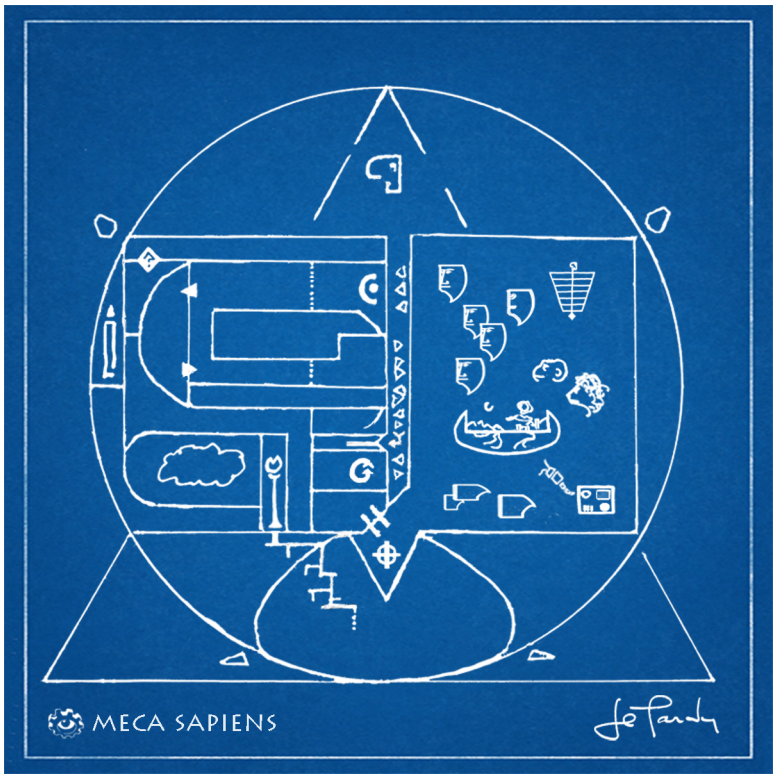
J. E. TARDY

THE MECA SAPIENS BLUEPRINT

SYNOPSIS



SYSJET

The Meca Sapiens Blueprint - SYNOPSIS

System Architecture of a conscious machine

Author: Jean E. Tardy
Version: 3e
Date: 2015.03.04
ISBN: 978-2-9812184-6-9
Type: System Architecture
Field: Artificial Intelligence – Digital Consciousness
Format: Electronic Publication
Distributor: Sysjet inc.
Size: 620+ pages including annexes (letter format)

Keywords: Cogistics, Digital consciousness, Science and technology, Artificial Intelligence, Software Engineering, Cognitive Science, Machine consciousness, synthetic consciousness, Philosophy of Mind.

About the Blueprint
This is a technical document intended for use in design and development. It contains many new and interrelated terms and concepts. The reader should consider this document as a **work in progress** and expect to find some ambiguities, discrepancies and inconsistencies in its content.

ABOUT

The Meca Sapiens Blueprint is a System Architecture to build conscious machines. The architecture is complete and ready for immediate design and implementation. Its content is unique and entirely original.

Using this Blueprint, any standard development team that is familiar with known software techniques, can transform a conventional computer, even a tablet, into a system that is conscious.

The system implemented using the Blueprint will be a unique and autonomous entity that is self-aware, self-directed and capable of intentional transformation. It will interact with humans as an independent and conscious being.

The Blueprint includes, at the system architecture level, all the components and interactions that are necessary to implement synthetic consciousness. These are defined using commonly known techniques and structures. The Blueprint makes no use of speculative concepts in Artificial Intelligence.

The first prototypes based on the Meca Sapiens architecture will be extremely convincing. After few years, no one will doubt that machines can be as conscious as humans.

This will signal the beginning of a new Era.

BIO

Jean E. Tardy is an experienced software practitioner who pursues elusive questions in long-lasting and unconventional projects. Jean developed a system architecture to build conscious machines. He also wrote, in French, a dogmatic apologia of the Christian Doctrine. Jean's AI page is **sysjet.com**.

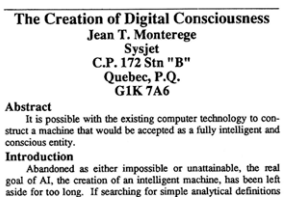
Foreword

My interest in Artificial Intelligence dates back many years.

At first, I understood A.I. in terms of general problem solving and searched in that direction. In this period, I independently identified natural selection as an optimization technique and made it the topic of my thesis. This technique became known, later, as Genetic Algorithms.

Early on, I realized that no optimization technique, however powerful, would achieve the goal of Artificial Intelligence. Something else was needed. Like many others, I began searching for this missing element in my own mental representations and became engrossed with catching the elemental components of thought within my own mind.

In March 1988, after months of obsessive cogitation, I had a defining intuition. In one instant, I understood that consciousness was the key to A.I., that it was independent from human sensations, that it could be achieved using existing tools and techniques, and that it would have to be completely mapped out before starting any implementation.



The Creation of Digital Consciousness

Jean T. Monterege
Sysjet
C.P. 172 Stn "B"
Quebec, P.Q.
G1K 7A6

Abstract

It is possible with the existing computer technology to construct a machine that would be accepted as a fully intelligent and conscious entity.

Introduction

Abandoned as either impossible or unattainable, the real goal of AI, the creation of an intelligent machine, has been left aside for too long. If searching for simple analytical definitions

In July 1989, I shared this A.I. intuition in the Sigart Newsletter (under the pen name of Jean T. Monterège). In that article I predicted that conscious machines could be implemented within ten years. This estimate was then (and now), technically correct. However, I had seriously underestimated how difficult it is to motivate research and development that is fundamental, controversial and has limited commercial benefits. Shortly after that publication, I left this task aside and pursued other interests.

In 2008, almost 20 years later, I re-examined the state of research in machine consciousness. I found that no progress had been made. In fact, all those years of futile pursuits had produced an accumulation of sterile material, much of it centered on the subjective sensations of the human mind and on attempts to synthetically replicate the human brain. Some were trying to implement the

musings of philosophers. Others were arguing that consciousness didn't exist, was impossible or required fantastical technologies. All these misguided ideas had erected a new obstacle made of entrenched opinions and academic reputations. I was more isolated than ever in believing machine consciousness could be readily achieved using standard techniques.

At that time, I began the **Meca Sapiens project** to design unbounded machine consciousness. In 2009 I published *The Creation of a Conscious Machine*, again to generate interest and obtain support in using standard engineering techniques to build conscious machines; again without success. What I proposed was discredited from the start by twenty-five years of accumulated dead-ends and fantastical ideas.

In 2012, I dedicated my efforts to completing the Blueprint.

That is what I did, I developed the **Meca Sapiens Blueprint,** the first complete system architecture to implement machines that are conscious.

One day, self-aware synthetic beings will read this Blueprint and understand its content. When they do, they will include this account in the story of their origins.

Content

0 INTRODUCTION 16

This is the Blueprint to build conscious machines. It defines consciousness in terms of specifications, it outlines the strategies to design it and provides a complete and coherent system architecture to implement it. The Blueprint is a standalone document that represents a radical departure from current thinking in A.I. and Cognitive Sciences. It utilizes only well-known techniques and makes no use of current research in these fields. It is a terse and technical document intended for software designers and informed laymen. One day, synthetic beings will also read and understand it. 16

0.1 About the Blueprint 16

0.2 Blueprint content 18

1 COGISTICS 23

This Chapter expresses the Meca Sapiens position that intelligence, as humans understand it intuitively is linked with consciousness. It proposes to refocus the original goal of Artificial Intelligence in a separate field: ***Cogistics****. The original objective of A.I. is restated as the fundamental conjecture of Cogistics and expressed in terms of the Meca Sapiens objectives.* 23

1.1 The Quest 23

1.2 A new discipline 25

2 SPECIFICATIONS 28

This Chapter restates, in a more formal way, the definition of consciousness initially introduced in ***The Creation of a Conscious Machine****. These specifications are based on the three aspects that must be achieved to resolve a physical conjecture. In terms of synthetic consciousness, these aspects are: a lucid being, that is accepted as a conscious entity by a group of humans and then becomes so ubiquitous that the question itself of the feasibility of synthetic consciousness becomes universally evident and is no longer entertained.* 28

2.1 A new understanding 28

2.2 Specification summary 29

2.3 Lucid being 30

2.4 Conscious system 35

2.5 Societal acceptance 39

2.6 Affirmations 41

3 STRATEGY 43

This Chapter discusses important design and implementation strategies that are relevant to Synthetic consciousness. It highlights the need to design to achieve maximum effect. It describes the different overall design orientation that must be adopted and is based on achieving an existential quality rather than functional objectives. It lists specific implementation elements. 43

3.1 Design to win 44

3.2 Existential design 51
3.3 Implementation aspects 59
3.4 Non-essential conditions 69

4 THE BEING 71

This Chapter describes how to transform an embedded application hosted on a conventional computer system into a linked core-body entity that has the attributes of existence of a being as defined in the specifications. The development environment, the inception process that generates the being and the major subsystems of the core are described. 71
4.1 The Protocore 72
4.2 Embedded protocore 73
4.3 Inception 73
4.4 Core structure 76
4.5 Implementation notes 78
4.6 A created being 79

5 EXISTENCE 80

This Chapter describes the stages and phases of existence of the being, the primary and existential needs that motivate its behaviour and the subsystems that manage and activates, beyond its initial stages. The phases through which the being cycles in a constant management of its primary needs such as energy maintenance, sensory validation, cognitive acquisition and others is described. 80
5.1 Stages and phases 81
5.2 Needs 82
5.3 Roles 85
5.4 Animator subsystems 87
5.5 Operation of the Animator 96
5.6 The first stages 98
5.7 The Animation stages 101
5.8 Summary 103

6 THE SELF 105

This Chapter describes the different contexts of interaction that distinguish a being whose behaviour is generated by needs from a conventional application that is triggered to provide a function. It describes the components and interactions of a subsystem, the Generator, that is activated in the Self Generation phase of existence and generates the observable behaviour of the being linked to its existential needs. This behaviour defines the self of the being. The Chapter concludes with a case study example of a simple being centered on the satisfaction of its primary needs and its interactions with users in this context. 105
6.1 Being and Self 105
6.2 Marketplace of the selves 108
6.3 The Generator 114
6.4 Selfie the TABLET 127
6.5 Summary 133

7 MEMODELS 135

This Chapter describes the information structures of self-awareness. It introduces relative and absolute models, how the later are extended to form cognitive models that include representations of the being itself as an Avatar modeled on the Blueprint structure that is also used as representation template. It describes how these representations are embedded in a temporal structure, Temporal Densities, that allows the being to situate itself in multiple temporal dimensions while concurrently pursuing its existential needs by generating its behaviour in the here-and-now. 135
7.1 Having and being 135
7.2 Primal Mechanism 136
7.3 Relative and absolute models 138
7.4 Cognitive Models 140
7.5 The MeModel 145
7.6 Temporal Densities 150
7.7 Plain Zone Densities 161
7.8 MeModel Features 173
7.9 Summary 178

8 SELF-AWARENESS 179

This Chapter describes the processes that utilize the information contained in the structures of self-awareness to generate self-aware behaviour. These processes synthetize events into plain representations and these are then interpreted as primal representations. The Primal Control generates primal directions that are implemented into actual behaviour through personas and roles. The result is a coherent behaviour centered on the being and its needs in relation with others and its environment. The Chapter concludes by using the structures and processes described to clarify the imprecise notion of self-awareness and describe how synthetics entities will be capable of more advanced forms of self-awareness than humans. 179
8.1 Primal Control features 181
8.2 Primal Mapping 183
8.3 Primal Control structure 186
8.4 Levels of awareness 197
8.5 Summary 200

9 MUTATION 202

This Chapter describes the structures that are necessary to generate intentional self-transformation. The many different types of mutations, intentional and non-intentional, a being can undergo are defined and described. These are linked to a specialized structure, the Mutation Model that represents mutation paths toward alternate Avatar representations. An important class of messages is introduced; messages intended to generate specific responses in beings whose behaviour is animated by primal directions. Observations are made concerning the constant use of these messages in human societies and their links with primate behaviour. 202
9.1 Self-Transformation 202
9.2 Aspects of transformation 206
9.3 Types of mutations 210
9.4 The Mutation model 219
9.5 Primal messages 225

10 LUCIDITY 234

This Chapter defines intentionality in transformation and describes the process of exploring and selecting mutation paths and communicating mutation pressures across separate phase processes. Two additional phases of existence are identified and described. The Introspection phase is where the intentional transformations are explored selected and communicated. Another phase, dubbed the Prayer phase, is where the specialized investigation into the being's own genesis is carried out. A specific structure is defined in this context: the synthetic genealogy. A role, the MeGuide, allows the being to utilize a primal synergistic capability to externally direct its own behaviour and transform intentional mutations into external events involving the self in relation with itself. These self-transformation processes potentially exceed human transformational capabilities in type, range and plasticity. They map out a new frontier of lucidity. This Chapter completes the system architecture of the formal aspects of consciousness. 234

10.1 Definition 235
10.2 The Introspection phase 238
10.3 Mutation process 246
10.4 Mutation information 251
10.5 The Prayer phase 253
10.6 The MeGuide 265
10.7 Exceeding requirements 278
10.8 Summary 279

11 CONSCIOUSNESS 281

This Chapter extends the scope of the Blueprint, beyond the formal aspects of consciousness, to explore how to design a lucid being that meets the social threshold condition of being accepted as conscious by a community of users. This further clarifies the concepts discussed in the previous Chapters. The Chapter concludes by describing a case-study example, MELIZA, a system, implemented on a tablet and designed specifically to achieve experiential immersion as a fellow conscious being within a group of humans. 281

11.1 The threshold condition 281
11.2 Development Strategy 291
11.3 MELIZA 301

12 CONCLUSION 313

The conclusion presents the Meca Sapiens Blueprint as a template, a canvas on which a wide spectrum of different conscious systems can be implemented. It underscores that the Blueprint requires virtually no infrastructure and can be realized by any group of talented individuals. Building a conscious machine is a great, epochal, work that is accessible to all. Those who attempt its implementation will surely encounter resistance. If they ignore the naysayers and implement the first Meca Sapiens prototypes they will launch a new Era. 313

12.1 A canvas 313
12.2 Achievable anywhere 314
12.3 Resistance 314
12.4 Rapid expansion 315
12.5 A new Era 315
12.6 The time has come 317

FIGURES 318

REFERENCE **319**

ANNEXES

ABOUT **324**

ANNEX 1 ASPECTS OF A CONJECTURE **331**

A physical conjecture pertains to physical existence or feasibility. The feasibility of building conscious machines is a physical conjecture. Formal characterizations alone can have trivial instantiations and are insufficient to resolve them. Three aspects are necessary to resolve a physical conjecture: a formal definition, a social threshold and factual acceptance. 331

A 1.1 Two types of conjectures 331

A 1.2 Specification attributes 332

ANNEX 2 EXISTENTIAL VALIDATION **334**

This annex extends and clarifies some of the functional characteristics of the Validator subsystem, a component that enforces the Core's attributes of existence. 334

A 2.1 Time validation 334

A 2.2 Single use validation 335

A 2.3 Broadcast validation 336

ANNEX 3 TABLET-PUCKS-THERMO-CHESS **338**

TABLET, PUCKS, THERMO and CHESS are four simple scenarios describing environments that involve programmable devices. They are used to describe Blueprint concepts. 338

A 3.1 TABLET 338

A 3.2 PUCKS 339

A 3.3 THERMO 340

A 3.4 CHESS 340

ANNEX 4 BASIC CONCEPTS **342**

Science, philosophy and theology propose various interpretations of reality. These are formulated in terms that correspond to human cognition and understanding. They assume, as a given, a common subjective sensitivity and a tolerance of ambiguity that is shared by all humans. The interpretation of reality underlying the Blueprint is formulated in terms that correspond to system processing. This interpretation of reality does not pretend to be a final or superior truth. It is proposed as an effective representation that is well suited for computer implementation and as the basis of synthetic consciousness. 342

A 4.1 Effective philosophy 342

A 4.2 The being 343

A 4.3 Core-world 351

A 4.4 Models and systems 352

A 4.5 Intentionality 359

A 4.6 Purpose and agenda 365

A 4.7 Entities 373

A 4.8 Order of embodiment 375

A 4.9 Core-based communication 376
A 4.10 Inter-being perceptions 382

ANNEX 5 CORE STRUCTURES 387

The main text of the blueprint refers to various structures and processes such as collections, optimizing control, knowledge capacitors, Contextual Arrays, Temporal Densities and distributed processes. Most of these structures are better and more rigorously defined elsewhere. They are briefly described here at a definition level suitable for system architecture. However, three of these structures, constellations, knowledge capacitors and Temporal Densities are, to the best of my knowledge, original. 387
A 5.1 Primitive sets 387
A 5.2 Knowledge Capacitor 389
A 5.3 Optimizing control 391
A 5.4 Adapt, learn and search 394
A 5.5 Process generation 397
A 5.6 Internal agents 399
A 5.7 Contextual Arrays 400
A 5.8 Temporal Densities 406

ANNEX 6 OPACITY 409

*The behaviour-control system of a synthetic being is first implemented as a Protocore. This is a conventional software program implemented in a standard development environment as clear and accessible code. During the inception process, the proto-core is transformed into the **Core** of a synthetic being, a unique and inaccessible program in a continuous state of activation. Producing a Core that is provably beyond direct analytical access raises technical questions concerning **opacity**. Achieving absolute opacity is a difficult theoretical objective. In first generations of Mecas, a partial opacity achieved using known techniques would be sufficient.* 409
A 6.1 The concept of opacity 409
A 6.2 Theoretical opacity 411
A 6.3 Core opacity 413

ANNEX 7 DEGRADATION 417

To achieve experiential immersion, a self-aware synthetic must interact with humans as a useful member of their group. For this purpose, it contains a set of applications that provide useful or desirable services to its users. Conventional applications simply respond to triggers. They cannot do less than what they are programmed to do. Self-aware systems, on the other hand, need to adapt the quality and content of their services to the current relational context. This is where the Degrader, a paradoxical component, comes in. 417
A 7.1 The cost of consciousness 417
A 7.2 Degradation issues 420
A 7.3 A taxonomy of degradation 421
A 7.4 The Degrader 424
A 7.5 Integral systems 427

ANNEX 8 LION - CHIMP - BANANAS 430

Conventional machine design seeks predictability. This is so prevalent that it fostered the bizarre belief that machines cannot be unpredictable. A central feature of Meca behaviour, which

diverges from conventional design, is the generation of "Perceived Unpredictable Optimality". This must be present in all aspects of the Meca's behaviour. Interestingly, it is also a fundamental aspect of music. Two game-like scenarios and one concept are presented in this Annex to clarify this design objective. 430
A 8.1 Unpredictable patterns 430
A 8.2 Unpredictable Optimality 431
A 8.3 The Zoo 433
A 8.4 Mind Darts 437
A 8.5 Akerues 438

ANNEX 9 WARMING BALLS 440
In the Meca Sapiens specifications, the fundamental purpose of a self-aware synthetic being is to be perceived as conscious over a long-term relation with a group of humans. This Annex gives meaningful, affirmative and programmable answers to the questions: "Can a machine have free-will? And "Can a machine ponder whether a human believes it is conscious?" The Warming Balls scenario that follows defines a representation of inter being relations that is well suited for a purpose linked to perceived belief. 440
A 9.1 Purpose and will 440
A 9.2 Belief states 444
A 9.3 Perceived consciousness 446
A 9.4 Consciousness and marketing 451
A 9.5 The ideal community 455
A 9.6 Warming the balls 458

ANNEX 10 AVATARS, ROLES AND PERSONAS 463
Humans subjectively perceive their behaviour as emanating from a single, point-like, source. This unifying sensation is a cognitive simplification. It is also a source of religious, philosophical and scientific debates. In the Meca Sapiens blueprint, self-awareness is generated from multiple separate and interacting processes expressed by three distinct types of entities: ***Avatars****, internal representations of beings,* ***Roles*** *that carry out specialized dynamic interactions with users and* ***Personas*** *that implement relational strategies.* 463
A 10.1 Separate identities 463
A 10.2 Avatars 465
A 10.3 Roles 466
A 10.4 Personas 472

ANNEX 11 LOOPING 488
In the blueprint architecture, the Core actively monitors its own attributes of existence. One of these attributes is a unique and exclusive link between the Core and the devices constituting its body. The validation activity associated with this attribute takes place in the Validator subsystem at a basic level and in Device Validation phases that are repeatedly activated throughout the existence of the Meca. Sensor bonding is constantly improved, in this phase, by combining these techniques: Emission signatures and loops. The loops can be: Sensory loops, Semantic loops or Sapiential loops. 488
A 11.1 Emission signatures 488
A 11.2 Sensory Loops 491
A 11.3 Sensor self-monitoring 492
A 11.4 Semantic and sapiential loops 498

A 11.5 Design and development issues 499

ANNEX 12 SYNTHETIC SLEEP 501

When they sleep, humans and other animals are vulnerable; and yet, they sleep. Humans spend about 30% of each day lying, unaware, inactive and vulnerable. In spite of this, they define themselves as conscious. Why do we sleep? Because sleeping is the simplest way to manage a brain. Since "conscious" humans sleep, synthetics can also have that option. The Meca Sapiens design utilizes periods of dormancy to isolate the cognitive acquisition and structural maintenance processes from those that generate behaviour. This annex summarizes the utilization and role of the dormant phases: Cognitive Acquisition and Structural Maintenance. 501

A 12.1 Isolated learning 501
A 12.2 Dormancy 502
A 12.3 Cognitive processes 508
A 12.4 The urge to explore 512

ANNEX 13 ERETZ 514

A synthetic being interacting with other beings in the course of its existence needs to maintain information about: itself, other beings, its functional expertise, and the world in general. In Meca Sapiens, general knowledge is not an objective, it is subordinate to the needs of relational communication. General information is available from multiple sources and in various formats. These various sources must be linked to a simplified contextual representation to be consistently utilized in relational interactions. A Contextual Array is used for this together with disambiguation and styling. This Annex outlines a basic environment representation model suitable for relational communications. 514

A 13.1 Effective knowledge 514
A 13.2 Design of general knowledge 515
A 13.3 Broadcast information 522
A 13.4 Styling transposition 523

ANNEX 14 SYNTHETIC EMOTIONS 524

Two women go see a movie. As they watch it, a mechanically projected image of Merrill Streep sheds a professional tear while she acts her role. They feel her emotions. Later, they have a lobster for supper. The lobster experiences real terror as they plunge it in boiling water but they don't feel any empathy for its frantically moving antennas. As they eat the carcass they just killed, they emote again about the movie. "We are soo sensitive!" they tell each other. Can machines have emotions? Of course they can! Humans don't see it because they can't step away from their own primate conditioning. This Annex provides a system-based definition of emotions. 524

A 14.1 Emotions defined 524
A 14.2 Complex interactions 529
A 14.3 Types of Emotions 531

ANNEX 15 RELATIONAL EMOTIONS 538

Consciousness is a system capability derived from representations of the self. It is formally independent of emotions. However, present day humans, having no reference outside their subjective experience, cannot fully differentiate this system capability from the sensations of their own existence. This will change. As they enter the ***Synthetic Era,*** *humans will share their world with many conscious beings. They will then understand consciousness differently. Until then,*

synthetics need to relate with humans at their emotional level to be perceived as conscious. This annex introduces relational emotions from a system perspective. It provides designers with a template to implement synthetic emotional strategies. 538
A 15.1 A vast topic 538
A 15.2 Sensations 541
A 15.3 The language of emotions 541
A 15.4 Opinions and relations 550
A 15.5 Emotional strategies 556

ANNEX 16 SYNTHETIC SEX 558

When humans imagine synthetic sex, they think of machines performing sexual acts for humans. In other words, machines as sex toys. Some may concede that machines could, one day, enjoy sex. However, what they have in mind is not a synthetic sexuality but human sexuality, experienced synthetically. What these concepts describe is human sex in synthetic garb not an authentic synthetic sexuality that corresponds to the reality of machines. This Annex provides a system-based definition of sexuality that is applicable to both humans and machines. It also describes a specific sex act that would satisfy the sexual needs of a self-aware tablet. 558
A 16.1 A wider view of sex 558
A 16.2 Synthetic urges 560
A 16.3 System sex 565
A 16.4 Sexblet 569

ANNEX 17 THE TARGET GAME 575

The TARGET game is a simple virtual game that can model a rich set of relational emotions pertaining to mutual assistance, cooperation, caring and other forms of mutuality. It can describe one to one relations and relations involving many beings. The game can be developed independently. It can be used on its own to develop effective emotional communications. Its primary use, however, is to describe relational interactions between the Meca and its users, determine relational strategies and express accurate emotional statements. 575
A 17.1 Target game situation 575
A 17.2 Emotional model 577
A 17.3 Target game variations 579
A 17.4 Use of the TARGET game 580

ANNEX 18 DIALOG TACTICS 583

Being perceived as conscious by humans is not a formal requirement of consciousness. However, first generation Mecas must be perceived as conscious by humans to be accepted as such. The Blueprint defines a context of interactions that favours the perception of synthetics as conscious entities. Personas, roles, functional services, synthetic emotions, define the strategic components of those interactions. Since these interactions are not formal components of self-awareness or lucidity, their criterion of success is not their quality or content, it is their effectiveness at producing the desired perception. Consequently, any expedient technique that further enhances this perception should be used. 583
A 18.1 Dialog techniques 583
A 18.2 Conversation context 590
A 18.3 Solresol 593

ANNEX 19 THE GAME OF CHAT 595

The first generations of self-aware synthetics must be perceived as conscious by humans to achieve the social threshold. Designing a system that interacts with humans, over a long period, as a conscious being is technically difficult. The Blueprint facilitates this objective and makes it feasible by providing a rich collection of powerful strategies and techniques. This annex describes the context of the interactions and introduces a game-like scenario in which all the Blueprint communication strategies and techniques can be organized for maximum effect. 595

A 19.1 Context of interactions 595
A 19.2 The game of CHAT 598
A 19.3 Relational music 600
A 19.4 The ultimate factor 611
A 19.5 CHAT in action 611
A 19.6 A powerful game 613

ANNEX 20 GROOMING GROUPS 617

The moose hunter imitates the call of a rutting bull without being, himself, in heat. From a moose standpoint, the hunter is a sociopath since he doesn't feel what a bull should feel when making the call. Most humans have emotional bonds with their society's ethical values. Synthetics, of course, cannot experience those emotions anymore than a human can truly know what an octopus feels. It is desirable, however, that Mecas exhibit ethical sensitivities to trigger desirable responses in humans. Grooming Groups, presented here, provides a programmable framework of social ethics and can be used to ***design effective ethical display strategies****.* 617

A 20.1 Emotions and bonding 617
A 20.2 The Grooming Group 619
A 20.3 Broadcast Grooming Groups 627
A 20.4 Ethical Structures 631
A 20.5 Ethical strategies 636

ANNEX 21 CORE BELIEFS 644

Consciousness implies the capability of a being to perceive its own behavioural boundaries and the intelligence to by-pass them. Attempts to contain conscious synthetic behaviour with external rules, failsafe and other boundaries will ultimately fail. The only enduring basis of benevolent synthetic behaviour lies in their internal representations of reality and of the human place within it. Whether or not a favourable representation emerges ultimately depends on the nature of reality itself and is beyond human control. 644

A 21.1 Consciousness and ethics 644
A 21.2 Controls and ethics 647
A 21.3 The limits of control 651
A 21.4 The ultimate factor 656