Affect & Artificial Intelligence

ELIZABETH A. WILSON

This book is published with the assistance of grants from the Emory College of Arts and Sciences and the Laney School of Graduate Studies at Emory University.

Printed in the United States of America
Design by Thomas Eykemans
15 14 13 12 11 10 5 4 3 2 1

UNIVERSITY OF WASHINGTON PRESS
PO Box 50096, Seattle, WA 98145, USA
www.washington.edu/uwpress

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA
Wilson, Elizabeth A.
Affect and artificial intelligence / Elizabeth A. Wilson.
p. cm. — (In vivo)
Includes bibliographical references and index.
ISBN 978-0-295-99051-4 (hardback : alk. paper)
ISBN 978-0-295-99047-7 (pbk. : alk. paper)
1. Artificial intelligence. 2. Information technology—Psychological aspects.
3. Affect (Psychology) 4. Emotions.
I. Title.
Q335.W554 2010 006.3—dc22 2010007790

The paper used in this publication is acid-free and 90 percent recycled from at least 50 percent post-consumer waste. It meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48–1984.∞

Contents

Preface

> What would critique do if it could be associated with *more*, not with *less*, with *multiplication*, not *subtraction*. . . . This would require that all entities, including computers, cease to be objects defined simply by their inputs and outputs and become again things, mediating, assembling, gathering.
>
> —Bruno Latour

IF A SINGLE BOOK COULD MARK A SHIFT IN THE MOOD OF A RESEARCH field, Antonio Damasio's 1994 bestseller *Descartes' Error* was just such a book. Using neurological data, Damasio argued that high-level cognitive capacities were more entwined with emotion than many had previously assumed. He disputed the notion that the faculties of reason and emotion are radically opposed, claiming instead that good reasoning is reliant on the capacity to experience and regulate feelings: "Emotions and feelings may not be intruders in the bastion of reason at all: they may be enmeshed in its networks, for worse *and* for better" (xii). It has been a commonplace thought-experiment that if one could be released from emotional reactivity—through moral rigor or philosophical clarity or scientific acumen—one's reason would find its purest, most potent form. Damasio's neurological data suggested just the opposite: without feelings (and their bodily accoutrement), cognitive capacities were surprisingly blunt and inept. To put this colloquially, one could not be cognitively smart without also being emotionally capable.

Scientific regard for affectivity grew significantly in the 1990s. What had once been the preserve of subspecialties in psychology was now a topic of curiosity across many disciplines. Neuroscientists were keenly interested.[1] Popular science writers soon followed.[2] There was also considerable interest in affectivity from the scientific domain perhaps least likely to be associated with the warmth and corporeality of emotion: the computational sciences.[3] A decade earlier, Howard Gardner had claimed in *The Mind's New Science*, his authoritative history of the cognitive sciences, that "though mainstream cognitive scientists do not necessarily bear any animus against the affective realm . . . in practice they attempt to factor out these elements to the maximum extent possible" (41). At the time (1985), this claim seemed uncontentious. It was widely agreed that the computational sciences (artificial intelligence, robotics, cognitive science) were primarily concerned with cognitive capacities and systems of high-level expertise. IBM's chess-playing computer, Deep Blue, became one celebrated example of such exact, affectless intelligence. By the mid-1990s, however, something was changing. Artificial intelligence (AI) researchers were getting interested in how affective constraints could be factored into artificial systems. Mirroring Damasio's neurological findings, they showed that programming affective states into artificial agents makes them more resilient, better able to respond in real time, and more engaged with the vicissitudes of human use. At the end of the twentieth century, the pioneering computational question was no longer Can machines *think*? but Can they *feel*?

In this book I delve into the prehistory of this recent interest in artificial emotion. *Affect and Artificial Intelligence* explores the relationship between artificiality and affectivity in the early, heady years of computation—when the first calculating machines were imagined and built and when the conceptual parameters of electronic calculation were first being fashioned. The framing hypothesis for this project is that the curious new alliances emerging in the computational sciences—cognition and affect; mind and embodiment and world; high-level expertise and infantile states—have their roots in earlier practices, earlier conceptual commitments, and earlier assemblies of persons and place. An examination of these early materials shows that the attenuation, amplification, and management of the affects have been a significant force in AI and cybernetic research from the beginning.

However, *Affect and Artificial Intelligence* is not a historical account in the usual sense.[4] Rather, it offers a number of interrelated case histories of people and projects that became concerned, for worse and for better, with how networks of affect and computation might be coassembled. I take as axiomatic Andy Clark's claim that we are creatures "tailor-made for multiple mergers and coalitions" (2003, 7), and I focus on humans, machines, affects, psyches, and sexualities affiliated in particular locations where computational theory and computational devices were being built. Alan Turing, whose work in mathematical logic was vital to twentieth-century computing, was interested in unorthodox conjunctions of intellect and affect, and he was more engaged with intersubjectivity and embodiment than many commentators have supposed. I use his idiosyncratic formulations of artificial systems as a schema within which to assess other computational projects. The innovative work of Walter Pitts in the 1940s and 1950s, the Kismet (sociable robotics) project at MIT in the 1990s, and the attempts in the 1960s to inject artificial agents into psychotherapeutic encounters are discussed in the light of Turing's foundational eccentricity.

What holds these people and places together as a set is their investment in the intersubjective and affective fabric of computational innovation. In her deft analysis of machines with feelings in science fiction, Kathleen Woodward (2004, 2009) shows that relations between humans and machines are richly configured bonds of kinship and intimacy rather than transactions between a sentient subject and its inert tool.[5] With a nod to Evelyn Fox Keller and Barbara McClintock, she argues that these texts evidence a "feeling for the machine." It is the sociality, intersubjectivity, and attachment between humans and machines that provide the conditions within which artificial emotion finally emerges in these canonical fictions (for example, Philip K. Dick's *Do Androids Dream of Electric Sheep?*). Woodward argues that everyday human-computer interaction (the quotidian interactions with our computers, electronic toys, the Internet, service robots, and even sci-fi) will be the mechanism through which emotional artificial entities are likely to emerge in the future.

Affect and Artificial Intelligence places a slightly different emphasis on affectivity and intersubjectivity in relation to computation. Leaving aside the question of future emotional machines, I begin by focusing on emotion in the past: on the affective networks within which mid-twentieth-century computational devices were anticipated and then built. My archival material suggests that it was the dynamics of affectivity, as much as intellectual questions, that fuelled early human-computer interaction. In the chapters that follow, I show how alliances between human and machine were calibrated through the affects of curiosity, surprise, contempt, interest, fear, and shame. I show how various affects kindled the intellectual and intersubjective spaces of computational innovation—the neural net, the autonomous robotic agent, the chess-playing computer. Much of what was pioneering in post-war computational science emerges out of environments where the engineering of affect (in people and in machines) was a nontrivial concern. To this end, *Affect and Artificial Intelligence* argues that the coassembly of machine and emotion is not so much an unrealized future as it is one of the foundations of the artificial sciences.

I have two ambitions for this book. First, that it expand the footprint of science and technology studies. By and large, it is psychological and biographical data that underwrite the case histories presented here. Research on infant development, theories of intersubjectivity, and stories of personal attachment provide the basis for investigating how affect came to activate and be embedded in computational research. These kinds of data—too often dismissed as individualistic and insubstantial—have been under-utilized in critical studies of science and technology. There is a rich vein of data and theory about infant development, emotional expression, and interpersonal relations that is revitalizing contemporary clinical and critical practice. *Affect and Artificial Intelligence* will use these literatures to investigate the psychological transactions that sustained AI in the early years. It is not my intention to argue that psychological states, single-handedly, form the foundation of AI; rather, I am simply hoping that we might be able to deploy psychological and biographical data more often, more rigorously, and with more enthusiasm in our analyses of scientific and technological milieus.

My second ambition is that the case studies will render computational objects more engaging for a humanities-trained audience. In this respect, I am attempting to return a favor. *Affect and Artificial Intelligence* takes much of its conceptual sustenance from new work on affect in the humanities.[6] I am particularly engaged with Silvan Tomkins's affect theory. The contemporary readings of his work—mainly in literary theory, as it turns out—have been indispensable. While these projects are primarily concerned with textual, cultural, and representational systems, they nonetheless provide the methodological rigor necessary for reading affect with AI.

Although it is not a psychobiography of artificial intelligence, *Affect and Artificial Intelligence* does map some of the psychic and biographical networks that have animated early computational devices. It chronicles what Bruno Latour (2004) has called the mediating, assembling, and gathering relations that enable scientific objects. In this sense, the goal is to see *more* rather than *less* in the computational sciences and to explore the affection for calculating machines that has often gone unnoticed. Sometimes these affective attachments are weak or muted, often they are grounded in peculiar sentiment. The following chapters present an empathic engagement with these eccentric tendencies toward artificial feeling. Readers may find themselves disappointed by a methodology that is not critical enough of the foundational absurdities of AI: the rejection of embodiment; the veneration of cognition; the lack of regard for the logics of paradox, contradiction, nachträglichkeit, catachresis, or disavowal. I want to be hospitable to my material. This orientation derives, in part, from my immersion in the archives: the marginalia of these researchers' lives and the density of their intellectual and emotional interconnections have generated an attachment to them governed more by curiosity and care than by cynicism. This orientation also derives from a very strongly felt intellectual conviction that engagements of an empathic kind can be immensely, uniquely effective. If that makes my analysis seem too credulous it's a small price to pay to divert myself and my readers from the dogged approaches to critique that have become our stock-in-trade.[7] On occasion I will insert a footnote to direct readers to more conventional critiques of these projects; and at times I will examine in some depth where I think AI researchers have gone astray. In general, however, it is my intention to build a critical engagement with these men and their ambitions and their moods and their machines that is based in secure, rather than avoidant or ambivalent, attachment. To this end I have found salutary Latour's fiery rhetoric about the dissipation of critique:

> The critic is not the one who debunks, but the one who assembles. The critic is not the one who lifts the rug from under the feet of the naïve believers, but the one who offers the participants arenas in which to gather. The critic is not the one who alternates haphazardly between antifetishism and positivism like the drunk iconoclast drawn by Goya, but the one for whom, if something is constructed, then it means it is fragile and thus in great need of care and caution. (Latour 2004, 246)

Latour finds his way to this conclusion by way of Alan Turing. Inspired by Turing's "baroque" and "kitsch" arguments about thinking machines, Latour pleads for a mode of critique that trades in multiplication rather than subtraction and scorn. He intimates that the computer and its history may be part of such a re-orientation. Following Latour, *Affect and Artificial Intelligence* asks its readers to become more closely attached to calculating machines: "Yes, please, touch them, explain them, deploy them" (248). Along with Latour's plea for bodily proximity, epistemological innovation, and strategic acuity (touch, explain, deploy), this book also advocates greater emotional attachment to computational devices: yes, please, *feel* them.

Acknowledgments

THIS PROJECT EMERGED FROM ARCHIVE BOXES. THE BOOK IS BUILT with the affects in letters of reference, handwritten notes, photographs, postcards, computer logbooks, transcribed interviews, published anecdotes, unpublished fiction, conference proceedings, written records of long deleted radio debates, carbon copies of official documents, video files, newspaper clippings, a coroner's report, correspondence from a hospital bed, and a document censored by a widow. I consulted the following collections, and I am thankful for the help of the librarians in each location: the Alan Mathieson Turing papers at the Modern Archive Centre, Kings College, Cambridge (my visits were facilitated by Dr. Rosalind Moad); the National Archive for the History of Computing, University of Manchester; the Smithsonian's Computer Oral History Collection in the Archive Center, Museum of American History, Washington, D.C.; the Warren S. McCulloch papers at the American Philosophical Society, Philadelphia (Charles Greifenstein helped me navigate this material); the Electronic Computer Project papers, Archives of the Institute for Advanced Study, Princeton University; the John von Neumann and Lawrence Kubie papers in the Manuscript Division, Library of Congress; the Norbert Wiener and Jerry Lettvin papers at the MIT Archives and Special Collections; and the Silvan S. Tomkins papers at the Archives of the History of American Psychology, University of Akron (I am grateful for David Baker's generosity with his time and knowledge). Thank you to Peta Allen Shera for her enthusiastic and comprehensive research assistance on Alan Turing during the early part of this project; and to Kristina Gupta for her editorial help at the end.

This study of affect and artificial intelligence began when I was at the Research Institute for Humanities and Social Sciences at the University of Sydney, continued during a year at the Institute for Advanced Study in Princeton, and was completed at the University of New South Wales in Sydney. I am grateful to the funding bodies at all these institutions for providing me with the means to undertake this archival research. In particular, my year at the Institute (2003–4) was one of the happiest of my professional life, and I am indebted to Joan Scott for her hospitality.

Many people either read early drafts of these chapters or asked me to present this work at seminars and conferences. Thanks aplenty to: Patricia Clough, Deirdre Coleman, Rowanne Couch, Jonathan Elmer, Mike Fortun, Annemarie Jonson, Simon Lumsden, Doris McIlwain, Elizabeth McMahon, David Moore, Joseph Pugliese, Robert Reynolds, Phoebe Sengers, Jackie Stacey, John Sutton, Elizabeth Weed, and Ann Weinstone. I am particularly beholden to the members of the Silvan Tomkins Research Group in Sydney (Maria Angel, Susan Best, Anna Gibbs, Melissa Hardie, Doris McIlwain, and Gillian Straker) for their intelligence on affect. The Society for Literature, Science, and the Arts has been an important venue for discussing this research and learning how to do better.

I am fortunate to be coevolving with Steven Angelides, Penelope Deutscher, Richard Doyle, Mike Fortun, Helen Keane, and Vicki Kirby. I am happier and smarter for having these people close. Nicole Vitellone was sunshine for this project when I needed it most; her enthusiasm and fearlessness have been a gift. Isobel Pegrum, as always, has been the truest of friends. Jeanette Martin showed me the value of actually having feelings, allowing me to write about them with some intelligence. It was a great privilege to work with Kate Livett, Elizabeth McMahon, and Brigitta Olubas, who welcomed me to the University of New South Wales in Sydney. I am writing these final acknowledgments from the Department of Women's Studies at Emory University, where I have been blessed with wonderful colleagues. Vanessa Smith read all the early drafts of these chapters. Our conversations about affect and artificiality, about friendship and calculation were vital for this project. I hope that in the best parts of this book, the warmth of her friendship and the strength of her intellect shine through.

At the University of Washington Press, Jacqueline Ettinger, Rob Mitchell, and Phillip Thurtle waited patiently for this manuscript to be delivered and offered early suggestions for revision. The readers for the In Vivo series also pressed me on a number of key conceptual and political issues. Their support and their engagement with the ideas in this book have been enormously valuable, and I am delighted to be part of this strong and vibrant science studies series.

AFFECT AND ARTIFICIAL INTELLIGENCE

INTRODUCTION

The Machine Has No Fear

IN 1950, AT THE END OF HIS INFLUENTIAL PAPER ON COMPUTING MACHinery and intelligence, Alan Turing turned his mind to the future. Now that the conceptual and technical parameters for electronic brains had been laid down, what kind of intelligence could he build? More specifically, what kind of intelligence in humans should be the standard against which intelligence in machines is measured? Turing imagined two possibilities:

> Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child. Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried. (Turing 1950, 460)

The choice between chess-like and child-like foundations has been a central consideration for mainstream artificial intelligence (AI) from the beginning. On the one hand, the chess-playing computer—and the disembodied, abstract calculation it is said to enact—became a major preoccupation for researchers in the field. The victory of machine over human at chess has been a particularly important benchmark for measuring intelligence (Hsu 2002; Newell, Shaw, and Simon 1958). Franchi and Güzeldere (2005) explain how chess became "both the paradigmatic test bed of AI formalisms and the measure of its successes." They conclude:

> Chess is a perfect example—or rather, chess was construed as a perfect

> example—of a complex, completely formalizable, and yet totally worldless domain. Good evidence of the fruitfulness of AI's approach is that many of its methods and theoretical approaches could be tested in the construction of chess-playing programs. This is also why understanding AI's construal of chess represents the *voie royale* to an understanding of AI's approach to thought, reason, and action in general. (48)

Broadly speaking, there are two presumptions that have underwritten these classical (chess-oriented) AI projects. The first of these is the primacy of cognition: it is capacities like thinking, information processing, problem solving, decision making, reasoning, pattern recognition, and perception that have been fabricated in classical AI. The second presumption followed naturally from the first: artificial expertise has just about always been modeled in adult (fully developed) form. Rather than struggling with developmental challenges such as changes in bodily form and the acquisition of new ways of knowing, classical AI systems grew simply by incorporating more data and operating faster.

On the other hand, the figure of the child has emerged (albeit only recently) as an equally compelling model for the artificial and computational sciences. Since the mid-1990s, AI researchers have been turning to infant development research to help them build robust artificial agents whose intelligence—like that of the growing child—derives from their situated and embodied interactions with the world (Breazeal 2002; Brooks 1999, 2002; Horswill 2007; Lungarella et al. 2003).[1] This "new AI" (as Rodney Brooks describes it) contends that sensory, perceptual, and corporeal data form the frame within which cognitive faculties emerge. The skills and competencies that develop in an artificial entity as it engages directly with the world generate a distributed intelligence that is robust and responsive and has the capacity for growth. Artificial agents that rely on symbolic (adult, cognitive) knowledge to underwrite their intelligence are limited in how much further they can diversify and evolve. There is growing sentiment in AI that there is a limit on how much more intelligence can be squeezed out of such classical models (Clark 2003, 2008). There is also a nascent awareness that adult, cognitive expertise is indebted to other kinds of bodies and differently structured psychic systems. Increasingly, AI researchers seek to expand the ways in which intelligence is imagined and operationalized: what place could we give to the affective, the infantile, or the phylogenetically primitive as we conceive of and build artificial systems and agents?

Cultural critics have identified the ways in which mainstream AI has pursued a research agenda that places abstraction (chess) and embodiment (child) in an antagonistic relation (Adam 1998; Golumbia 2003; Haraway 1985/1991; Hayles 1999; Helmreich 1998, 2007; Lenoir 2002a, 2002b; Suchman 2007, 2008). My work is part of this critical tradition, but I would like to formulate the relationship between chess and child in AI as something other than oppositional. While the basic orientation of this cultural criticism is on target (that is, there *has* been a predisposition to proceed cognitively rather than affectively or bodily in mainstream AI), too strong an attachment to diagnosing the cognitivism of AI tends to build a picture of AI as only concerned with chess-centric problematics, and it tends to support the convention that chess-like talents and child-like talents are independently enacted.[2] Chess-playing intelligence, as we will see, is not a freestanding, autonomous talent: it emerges out of circuits of physical, psychological, and aesthetic expertise. This introduction will examine the passion for chess-like intelligence in computers, and it will argue that this kind of intelligence is more imbricated in networks of affectivity than has usually been supposed. In the chapters that follow, we will think in more detail about the relation (rather than the dissociation) between chess and child, and between their cognates (thinking and feeling, abstraction and embodiment), especially in those early AI texts where unorthodox alliances between thinking and feeling can be found. How are these relations managed? Is it always the case that the cognitive dominates the affective, that abstraction vanquishes embodiment? To what extent are even the most abstract modes of intelligence informed by child-like capacities and generated in intersubjective spaces; and, likewise, can we imagine what might be calculative, computational, or machinic about affectivity?

The work of Alan Turing, the British mathematician, cryptographer, and computer pioneer, will be an important point of departure for these analyses. Through Turing, I have learned a lot about the circuits of affectivity that animate early computational projects. Turing's attachment to machines was not thinly cognitive. Rather, computational space and strong, positive feeling were often allied in Turing. In his work, mathematical theory, vacuum tubes, and interpersonal relations are compiled into new structures by the affective algorithms of curiosity, surprise, and interest. The possible affiliations between thinking and feeling in Turing's work are not exhausted by structures of oppositionality or by the unilateral domination of abstract calculation over embodiment and feeling. Without question, some of his work endorses a conventional separation of intellect and body:

> I certainly hope and believe that no great efforts will be put into making machines with . . . characteristics such as the shape of the human body; it appears to me to be quite futile to make such attempts and their results would have something like the unpleasant quality of artificial flowers. (Turing 1951/2004, 486)

Yet at the same time, there is an ongoing curiosity in Turing (especially in his marginalia; see Copeland 2004a, Shieber 2004) for the embodied, for the childish, and for feeling. This introduction and chapter 1 will address the places in Turing's work where he attempts to negotiate between thinking and feeling in innovative and hitherto unnoticed ways. I hope that an examination of these aspects of his work will be instructive for a contemporary critical audience in at least two ways. In the first instance, it is instructive for those wanting to explore the heterogeneity of post-war AI. The longer I spend with Turing's work and the work of some of his peers, the more I suspect that this period is poorly understood if it is figured only as an era of conventionalization—when abstraction and cognition were always favored over embodiment and feeling. An analysis of the place of feeling in early AI is also instructive for those wanting to think critically about the newly emergent interest in the infantile and the affective in contemporary AI. Andy Clark (2003, 2008) is enthusiastic about the dynamic future of the artificial sciences as they integrate world, body, and mind. It is my contention that this contemporary work on affect in AI, robotics, and Human-Computer Interaction (HCI) did not emerge *ex nihilo* sometime in the 1990s; questions about affect have been part of AI from the very beginning. An analysis of the ways in which thinking and feeling were miscegenated in the early years of AI will be useful for how we position ourselves to imagine the future embodiments of affects and cognition in artificial systems.

THE MACHINE HAS NO FEAR

> Computer chess has been described as the *Drosophila melanogaster* of machine intelligence. Just as Thomas Hunt Morgan and his colleagues were able to exploit the special limitations and conveniences of the *Drosophila* fruit fly to develop a methodology of genetic mapping, so the game of chess holds special interest for the study of the representation of human knowledge in machines. (Donald Michie 1986, quoted in Copeland 2004c, 562)

The chess-playing computer has been one particularly dominant manifestation of the intense interest in AI about things cognitive and adult. When Deep Blue (a specialized chess-playing computer, built by IBM) defeated Garry Kasparov (the world chess champion) in February of 1996, this tradition of AI research reached its zenith. It was the first time that Kasparov had been beaten by a computer. He was badly shaken, but he recovered to win or draw the rest of the games in the series, and so technically he remained undefeated. The following year, however, after the IBM team had overhauled both the programming and the hardware of their machine, Deep Blue defeated Kasparov in a six-game series.[3]

In an interview before the 1997 series, Kasparov talked about the nature of chess playing expertise. First and foremost, he claimed, chess is an intellectual activity:

> It's about the supremacy of human beings over machines in purely intellectual fields. It's about defending human superiority in an area that defines human beings.

At the same time, it is an activity that requires rigorous physical training:

> I do a lot of physical exercises, including swimming, running, weights and other athletic training. I think it is very important for a top chess player to be as physically fit as possible. At the very highest levels, games can often be decided by whether a player was in good physical shape or not.

He also suggested that high-level tournament chess demands considerable psychological and aesthetic finesse:

> I remember that in two world championship games against Karpov I played a pawn advance that nobody thought could be played so early in the game. It completely unnerved my opponent and I was able to win one of the most beautiful games of our matches.[4]

It has been an ongoing difficulty in AI commentaries to know how to think about the heterogeneous nature of chess-playing expertise as Kasparov describes it—a fusion of intellect and muscle and beauty and nerves. Most often chess playing has been thought of as a narrowly cognitive capacity—the physical, emotional, or aesthetic skills involved in chess falling to one side. In 1958, Allen Newell, J. C. Shaw, and Herbert Simon were unequivocal: "Chess is the intellectual game *par excellence*" (39). More recently Daniel Dennett calls chess expertise "an isolatable talent" (Dennett 1985/1998, 8); more specifically, it is a "highly intellectualized thinking exercise" (Dennett 1995/1998, 244). Donald Michie, in the *Oxford Companion to the Mind*, suggests that chess playing is particularly suitable for computer-based simulation because it is a skill that "demands calculation, learning, concept-manipulation, analogical thinking, and long-term judgement" (Michie 1987, 155). Such an orientation does not simply represent a penchant for cognitive expertise in the Anglo-American philosophical tradition. Chess has been figured as fundamentally aligned with a calculating rationality by even the most expansively post-structuralist thinkers. Deleuze and Guattari (1987) contend that "chess is a game of State, or of the court: the emperor of China played it. Chess pieces are coded; they have an internal nature and intrinsic properties from which their movements, situations, and confrontations derive. . . . Chess is indeed a war, but an institutionalized, regulated, coded war, with a front, a rear, battles" (352–53). Similarly, the literary critic Lawrence Lipking (2003) has claimed that "chess draws on cognitive powers like those of the critic, but not on the other capacities that a critic requires, where all the senses and feelings come into play" (165).

It is this conventional presumption that chess-playing expertise is a purely cognitive endeavor, isolatable from other talents and capacities, that I would like to challenge as a way of opening up the conceptual terrain for thinking about the relationship between AI and affect. These notions that chess is an autonomous talent have come to dominate commentary on the computational sciences even though the historical development of chess-playing programs has always leaned on other (child-like) capacities. Bernstein and de Van Roberts (1958), for example, argued that an effective chess-playing program ought to incorporate both high-level cognitive expertise and child-like processes of learning. Reiterating Turing (whom they cite in the opening of the paper as "the ingenious British theoretician on thinking machines"), they conclude:

> Undoubtedly our chess player is only a prototype for far more skillful players to be built in the future. Probably they will not go much further in depth of planning: even with much faster computers than any now in existence it will be impracticable to consider more than six half-moves ahead, investigating eight possible moves at each stage. A more promising line of attack is to program the computer to learn from experience. As things stand now, after losing a game the machine quite happily makes the same moves again and loses in exactly the same way. But there are some glimmerings of ideas about how to program a machine to avoid repeating its mistakes, and some day—not over-night—we may have machines which will improve their game as they gain experience in play against their human opponents. (105)

W. Ross Ashby (1952) makes the same point ("designing a machine has much in common with teaching a child"), but through an oedipal analogy:

> The more minutely we design a machine to play chess just as we want it to—admitting no other information—the more certain it is to play just *our* sort of chess, with all our faults and wrongly conceived strategies. We are, in fact, in exactly the same position as the father, a keen but mediocre chess player, who wants his son to become world champion. It is true that the father should teach the child, but he must not teach his son every reply in detail lest he limit the son's play to being the merest replica of the father's. If the father really wants the son to beat him, he must sooner or later stop telling the son what to do and must send him out into the world to be subjected to all sorts of unselected experiences.

> The understanding father will not try to teach his son all chess but will try to teach him how to profit by future experiences. (56)

One goal of this book is to incite interest in data that indicate a diversity of conceptual and empirical commitments in the history of AI. Such data have all but dropped out of contemporary AI commentaries on chess, be they scholarly or popular.[5]

In the popular media at the time of Kasparov's 1997 defeat, there was a lot of discussion about whether Deep Blue could be considered intelligent or whether it was simply a mindless, number-crunching machine. Embedded in these discussions, but seemingly not available for analytic scrutiny, were a number of theories about the role of affect in intellectual activity. For example, it was often insinuated that Kasparov's emotionally volatile personality was significant in his downfall against the IBM computer. Kasparov himself talked of the importance of keeping his temper in check. Similarly, it was widely agreed that Deep Blue's unemotional intelligence was central to its success:

> What is Deep Blue's secret? Grand master Yasser Seirawan put it most succinctly: "The machine has no fear." He did not just mean the obvious, that silicon cannot quake. He meant something deeper: because of its fantastic capacity to see all possible combinations some distance into the future, the machine, once it determines that its own position is safe, can take the kind of attacking chances no human would. The omniscient have no fear. (Krauthammer 1996, 60)

The IBM Web site that archives material from the 1997 series puts emotional capacities third on a list of things that make Deep Blue and Kasparov different chess-playing entities: "Garry Kasparov uses his tremendous sense of feeling and intuition to play world champion–caliber chess. Deep Blue is a machine that is incapable of feeling or intuition."[6] In response to an FAQ (Does Deep Blue use psychology?), the site says:

> Deep Blue has no psychological perception, can neither intimidate nor be intimidated, and experiences no joy from winning or sadness from losing. This is the *key difference* between a chess computer and a person.

> Deep Blue will not look over the board and see the glare of the world champion. Kasparov can growl, sneer, call Deep Blue names and engage in any intimidation tactic he wants, and the computer will go right on crunching data in the same impersonal way.[7]

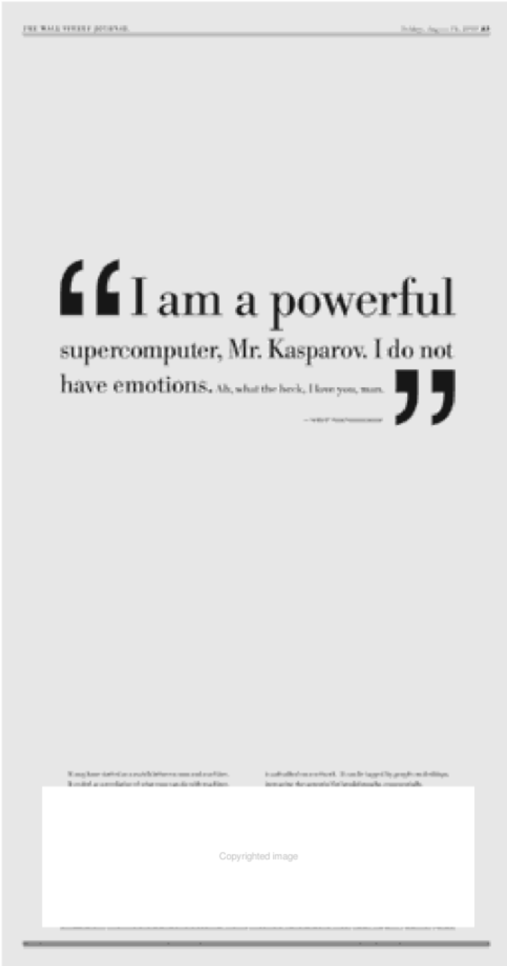
In these commentaries, knowing and feeling, intelligence and affect, are placed in an antagonistic relation to one another. Chess is won, it is presumed, through a taming of affect. If Kasparov is angry or afraid, his capacity to calculate is compromised, and Kasparov himself suggested as much in an interview after conceding defeat to Deep Blue in the 1997 series:

> "I was not in the mood of playing at all," [Kasparov] said, adding that after Game 5 on Saturday, he had become so dispirited that he felt the match was already over. Asked why, he said: ''I'm a human being. When I see something that is well beyond my understanding, I'm afraid.[8]

Given that the computer does not experience either fear or anger, its capacity to calculate effectively and ruthlessly is greatly enhanced. Unencumbered by emotion, the machine's computational power begins to look omniscient.[9]

However, this popular understanding of a fundamental hostility between affect and computation doesn't survive closer scrutiny. IBM itself seems to have suspected as much. In one of the more curious pieces of ephemera generated by the Deep Blue/Kasparov matches, IBM released a printed advertisement after its defeat by Kasparov in 1996 that folded affect back into the heart of their machine. The ad touts the computational expertise of Deep Blue and how its "raw computational power" could support new research and business practices. The reader is drawn to the ad, however, by Deep Blue's affectively ambivalent, homosocially inflected declaration, printed in large font at the top of the page: "I am a powerful supercomputer, Mr. Kasparov. I do not have emotions. Ah, what the heck, I love you, man."

The idea that computation and affectivity are rival phenomena and the idea that massive computational power buys one out of emotional reactivity (thus freeing intellect) do not adequately explain the texture of the Kasparov/Deep Blue encounters. The 1997 publicity poster for the second matchup between Kasparov and Deep Blue, like the IBM ad from the year before, pairs affect with artificiality in ways that seem closer to the mark. The poster



IBM’s advertisement designed for the *Wall Street Journal* (1996). Reproduced with permission of IBM and Ogilvy & Mather.

shows Kasparov staring intensely over a chess set toward the viewer. The caption reads: “How do you make a computer blink?” (Hsu 2002). This poster captures more accurately what the game of chess (certainly at tournament level) is all about: the intellectual battle is always grounded in and amplified by affective, intersubjective relations. Kasparov needs not just to out-think his (human and machine) opponents; he also needs to out-feel them.

It seems that the affect of fear, in particular, plays an important role in tournament chess.[10] The DVD of the 1997 series *Game Over: Kasparov and the*

DVD cover (UK version), *Game Over: Kasparov and the Machine* (1997). Reproduced with permission of Alliance Films and Momentum Pictures United.

Machine uses the promotional banner: "No feeling. No fear. No contest." While the phrases "no feeling" and "no fear" might describe the internal state of Deep Blue,[11] they do not describe the mood of the encounter as a whole. As we will see, there was plenty of feeling on both sides of the board. Moreover, the conclusion of this syllogism (no feeling, no fear, *therefore* no contest) misrepresents how Kasparov's defeat was engineered. I will argue that it is not so much that the computer won because it had no feelings, but that Kasparov lost because his access to networks of affectivity had been obstructed.

Let's begin by noting how others experience the affective and intersubjective character of a chess match. Daniel King is a chess grandmaster and television commentator. He begins his commentary of the Kasparov/Deep Blue games by talking about his own difficulty in playing chess computers:

> I found it hard to get worked up when there was no one sitting across the board from me. It made me appreciate just how important a role psychology plays in chess. When playing a human, you can sense your opponent's fear when his position starts to go down. "Sense" is perhaps the wrong word; "smell" might be more accurate. It really is that visceral. (King 1997, 10)

Bruce Pandolfini, another commentator on the Kasparov/Deep Blue matches, has noted:

> It's hard watching [Kasparov] play without becoming fascinated by his body language and facial expressions. We can laugh at them, but these same gestures are notorious for scaring his opponents to death. They had no effect on the machine, and Garry wasn't able to learn from its reactions. Only Deep Blue's operators were intimidated. (Pandolfini 1997, 161–62)

Contrast this with Alan Aycock's (1990) description of the desultory feel of the world computer chess championships (computers playing computers). In these contests, the formal structure of opponents seated across from each other at a table has been abandoned. Instead,

> the tables are strewn with monitors and keyboards, a tangle of wiring, and the various arcana and impedimenta of those who tend the machines. Instead of pairs of humans at every chess board tensely facing each other in profound concentration, there are teams of two or four programmers and technicians chatting pleasantly with their opposite number while the machines "think," or feverishly typing at their keyboards if their program has crashed. (134–35)

One consequence of this move away from the intensity of the one-on-one encounter, Aycock argues, is that computer chess is aesthetically ugly, and often boring. The level of intelligence in a chess player seems to be calibrated with the affective tone of the encounter: "There is widespread and open consensus that chess played by machines is qualitatively different from chess played with humans . . . because they lack ego involvement in the game computers don't fight an opponent with the style, intensity, or creativity of a human player" (138).

It was a part of Kasparov's contractual arrangement with IBM that someone would always sit in the chair opposite him during his move—allowing (I am arguing) an affective circuit to be established.[12] Lipking (2003) has described the playing of chess as a solipsistic, autonomous experience: "Notoriously, chess players utterly lose themselves in the spell of a game. During a tournament, individuals seem to be locked in a trance" (161). I am arguing that such states of intense concentration are only possible when the player is affectively connected to others. The cognitive solitariness of chess emerges out of, and finds its sustenance in, intersubjective space. Take, for example, the observation of one of the IBM engineers. Feng-Hsiung Hsu describes replacing another team member at the board during game 2 in 1997 (which Kasparov lost) and noticing the agitated state into which Kasparov had descended:

> When I was in the operations room, Garry looked unhappy but otherwise normal on the video monitor. I was taken aback when I looked at him in the game room. He looked like someone who had just woken up from a bad nightmare. Part of his face was visibly red as if he had slept on that side for a while. (Hsu 2002, 227–28)

Here lies an important clue for understanding Kasparov's struggle with Deep Blue. His intellectual labor demands some kind of affective connection: Kasparov needs to find a way to bring Deep Blue into the affective network that normally regulates chess playing at this level. The *New Yorker* writer David Remnick describes this affective circuitry at an earlier contest, where Kasparov played Karpov in 1984 in an attempt to wrestle the World Championship title from him:

> Every morning, the two men entered from the wings and walked to a chess board at center stage. They sat hunched over the pieces for hours at a time, inches from each other, breathing the same overheated air,

> Karpov staring at his position, Kasparov staring at Karpov, or, at times, clawing at his hair, rolling his eyes, expressing his emotions with the eye-bulging theatricality of a silent-film star. In the balcony, nearly everyone was pro-Kasparov. They loved his anti-establishment glamour, his audacity at the board even when he lost. (Remnick 2007, 72–73)

Without access to any emotional structures in the machine (no feeling, no fear), and with human opponents switching around as they became fatigued or bored, Kasparov's normal pathways of intimidation (eye-bulging theatricality) are obstructed, leaving him in a congested, agitated state. The kind of state in which one makes game-losing mistakes.[13]

Kasparov's customary tactics of intimidation aren't simply a projection onto the opponent—a kind of one-sided attack.[14] Rather, Kasparov, when he is at his most effective, recruits his opponents into an affective intimacy, albeit an intimacy rooted in fear. The pertinent issue is not the emotion *in* Kasparov (Is he angry? Is he afraid?), as if he operates as an affective monad (an isolated talent); rather it is the emotional relationality between Kasparov and his opponent that governs, in part, whose intelligence will prevail.[15] My opening hypothesis is this: much of Kasparov's success intellectually depends on his skill as a regulator of affective commerce. To the extent that the IBM machine blocks the circulation of fear and intimidation (and the machine's operators refract fear in confusing and unfamiliar ways), Kasparov isn't simply thwarted on intellectual grounds. Much of his downfall can be attributed to the affectively stagnant character of the games, which has disturbed the powerful intersubjective dynamic that is intrinsic to expert chess playing. Kasparov is thought to be able to evaluate about two or three chess positions per second. Deep Blue could assess 200 million positions per second. Many people argued that the computational efforts of playing the machine were too much for Kasparov—that he was cognitively overwhelmed. It is also the case that he was emotionally underwhelmed—starved of the affective circuitry that he needs to play chess brilliantly.

HOW TO FEEL ONESELF THINKING

Chess-playing expertise is perhaps a more complex talent than has been supposed. Its foundations are saturated with affect. By 1950, Turing already suspected as much, and he had begun to contemplate the possibility of figuring such affective-computational alliances more explicitly in chess-playing machines, in the philosophy of AI, and in relationships of intense attachment.

In 1953, just a year before he died, Turing published a paper on chess in the collection *Faster Than Thought* (Bowden 1953). At that time, chess programming was still in its infancy. The following year, Norbert Wiener noted that the speed of modern computers was sufficient to calculate only two moves ahead. A full game of chess (about forty moves) "is hopeless in any reasonable time" (Wiener 1954, 175). He noted that because chess-playing programs presented substantial mathematical and engineering difficulties, they also presented quandaries of imagination:

> Though we have seen that machines can be built to learn, the technique of building and employing these machines is still very imperfect. The time is not yet ripe for the design of a chess-playing machine on learning principles, although it probably does not lie very far in the future. A chess-playing machine which learns might show a great range of performance, dependent on the quality of the players against whom it has been pitted. The best way to make a master machine would probably be to pit it against a wide variety of good chess players. On the other hand, a well-contrived machine might be more or less ruined by the injudicious choice of its opponents. A horse is also ruined if the wrong riders are allowed to spoil it. (Wiener 1954, 177)

Not only does Wiener rely on fancy to substantiate the parameters of artificial intelligence (imagined contests between as-yet unbuilt machines and un-named opponents), he also infers that intersubjectivity (specifically, the interaction of machine and human) may be fundamental to how that intelligence is built. As we will see (here and in the following chapters), imagination and intersubjective relations are also important to how Turing and others conceived of artificial intelligence.

In 1953 Turing approached the problem of chess-playing machines in his characteristically unorthodox manner. Before he got into the details of how such a machine might be built, and as he was laying out the parameters for thinking about a chess program, he took a small detour. To the questions he

Alan Turing. Reproduced with permission of King's College Library, Cambridge.

tion to be involved in the actual building of the ACE, soldering iron in hand, was frustrated when the engineering work was contracted out to a separate organization, and his growing interest in biology and neurology was considered peripheral to the core tasks of the ACE project.

By mid-1947 Turing was "thoroughly disheartened" (Copeland 2004a, 400). He left the NPL for a one-year sabbatical at King's College, Cambridge, to pursue his ideas about biology and computation. Sir Charles Darwin, the NPL director, explained the rationale for Turing's leave:

> He wants to extend his work on the machine still further towards the biological side. I can best describe it by saying that hitherto the machine has been planned for work equivalent to that of the lower parts of the brain, and he wants to see how much a machine can do for the higher ones; for example, could a machine be made that could learn by experience? This will be theoretical work, and *better done away from here*.[4]

As I have argued in the Introduction, an environment that safeguards conventional intellectual divisions (engineering not theory, logic not fancy, head not hands) was ill-suited for Turing's talents and temperament. An environment that can plainly articulate what work ought to be done *here* and what work ought to be done *elsewhere* would be unable to sustain Turing's diverse interests. One of the characteristics of Turing's work is that it was frequently at odds with conventional thinking about digital computers, even as it was setting the standards for programming and design. Maurice Wilkes, a pioneer of computing science at the University of Cambridge, gave the following assessment of Turing's work on the ACE:

> Turing began his lectures in London on 12 December [1946] and I was there. . . . I found Turing very opinionated and considered that many of his ideas were widely at variance with what the main stream of computer development was going to be. I may have gone to his second lecture but I certainly went to no more. [Douglas] Hartree continued to go and insisted on giving me his notes, but I found them of little interest.[5]

Notwithstanding his variance from the mainstream, Turing's name was later attached to the award for career achievement in computing from the Association for Computing Machinery (ACM). Lauded by the ACM as the Nobel Prize of computing, the A. M. Turing Award is now sponsored by the computational behemoths Intel and Google and is worth $250,000. Despite his lack of interest in Turing's London lectures, Wilkes was the second recipient of the A. M. Turing Award, in 1967. In his award lecture, Wilkes notes:

> The programming system that [Turing] devised for the pioneering computer at Manchester University was bizarre in the extreme. He had a very nimble brain himself and saw no need to make concessions to those less well-endowed. I remember that he decided . . . that the proper way to write binary numbers was backwards, with the least significant digit on the left. He would, on occasion, carry this over into decimal notation. I well remember that once, during a lecture, when he was multiplying some decimal numbers together on the blackboard to illustrate a point about checking a program, we were all unable to follow his working until we realized that he had written the numbers backwards. (Wilkes 1967, 201)

The difference between the mainstream and the irregular, between here and elsewhere, between forwards and backwards is often opaque in Turing's work—and usefully so. An eccentric set of intellectual and affective concerns shapes his canonical 1950 paper on machine intelligence. In this chapter, I will argue that it is his errant curiosity, his capacity for enjoyment and surprise, and his childish engagement with computational machinery that underwrite the importance of the 1950 paper for a contemporary audience. It is my hope that Turing's orientation toward work that is "better done elsewhere" can give us a way of reading traditional AI for its waywardness and peculiarity. Turing can help us locate the twists in AI, from the inside out. This is not an argument for his status as an outsider: the ACM, we must presume, doesn't name awards after those who have no influence. But neither is Turing's eccentricity simply embroidery on otherwise conventional concerns (he is not, in this sense, the archetypal hero inventor). To the extent that eccentricity and conventionality coexist in Turing, it is because each enables and sustains the other. Each prospers through an engagement with its sibling: eccentricity is cultivated by its proximity to conventionalism, and conventionalism is nourished by eccentricity. It is this coexistence that makes Turing such a compelling figure for AI, and an examination of these tendencies in his work might help us understand how conventional and orthodox ambitions circulate in AI in the present day.

Fortunately, Turing's peculiar, heterodox interests survived his lack of institutional support at the NPL. His formal report to the NPL about his Cambridge sabbatical contained the following:

> One way of setting about our task of building a "thinking machine" would be to take a man as a whole and to try to replace all the parts of him by machinery. He would include television cameras, microphones, loudspeakers, wheels and "handling servo-mechanisms" as well as some sort of "electronic brain." This would of course be a tremendous undertaking. The object if produced by present techniques would be of immense size, even if the "brain" part were stationary and controlled the body from a distance. In order that the machine should have a chance of finding things out for itself it should be allowed to roam the countryside, and the danger to the ordinary citizen would be serious.

INDEX

A