

PERSPECTIVES IN BIOLOGY AND MEDICINE

Volume 42 · Number 4 · Summer 1999

THE ELUSIVE IMMUNE SELF: A CASE OF CATEGORY ERRORS

*ALFRED I. TAUBER**

Introduction

Immunology during the first half of the 20th century was preoccupied with the chemical questions of immune specificity, and the biological questions concerning immune identity were set aside [1]. But after World War II, transplantation and autoimmunity became increasingly relevant both to basic immunologists and clinicians. It was at this juncture that Sir Frank Macfarlane Burnet introduced the “self” into the immunological lexicon, and upon that metaphor erected a theory of immunological tolerance that was to dominate the field to this date [2, 3]. Most immunologists today define the discipline as the science of self/nonsel self discrimination, and Burnet’s clonal selection theory (CST) by which selfhood is understood, “has passed from the status of theory to that of paradigm” with only slight modification [4, p.15]. For those uncomfortable with such sweeping notions as “paradigms,” there is still a general consensus, as another textbook writes, that CST “is no longer a theory but a fact” [5, p. 335]. The Immune Self has indeed become dogma, and the “self” versus “other” axis has assumed the role of an operative thought style which organizes the entire discipline. I am most grateful to Moira Howes for initiating a philosophical discussion, perhaps a debate, with me over the nature of the immunological self [6]. This is a topic of great importance for current theoretical immunology—the science itself—as well as for various philosophical attempts at clarifying our understanding of personal identity. Let me offer a brief summary of the issues surrounding this topic by way of responding to Howes’ characterization of my work.

*Center for Philosophy and History of Science, Boston University, 745 Commonwealth Avenue, Boston, MA 02215.

© 1999 by The University of Chicago. All rights reserved.
0031-4982/99/4204-1113\$01.00

Today, the general view is that immunology, since its very beginnings during the last 25 years of the 19th century, was committed to discerning those mechanisms by which the “self” discriminates host elements from the foreign. In this view, the latter are destroyed by immune cells and their products, whereas the normal constituents of the animal are ignored. In other words, the identity of the host organism was given or assumed. But such neat divisions or boundaries were adopted, or at best were drawn with a certainty that remained problematic. Aside from competing theoretical formulations, there were early discrepancies accompanying the full embrace of a self/nonsel self discriminatory mode to explain immune function. Concomitant to demonstrating the beneficial effects of immunity, these same defensive functions were shown to be the cause of much of tissue damage. But inflammation only broadened the conceptual arena of immune mechanisms and was soon incorporated as the necessary untoward effects of the defense system. However, the immune system not only was found to inflict damage as the price of cleansing, it also was capable of apparently capricious assault on its host. So-called autoimmune reactions were described at the turn of the century and later determined as cause of autoimmune disease, but because the entire orientation of the science was to see immunity as a mediator of host defense, these findings were viewed as a pathological aberrancy. Arising from an unregulated killer system gone awry, autoimmunity, in this view, could hardly be regarded as part of an expected continuum of normal immune function. In each case, “ideal” immunity was the agent of the self, and although there might be inconsistencies in behavior regarding that mandate, the basic structure of immunology demanded articulation of a model of identification and the protection of organismal identity. In short, by the mid-20th century, a formal theory of self/nonsel self discrimination was articulated that would attempt to clearly demarcate the host organism and the foreign by an immune system that under normal conditions discerned the self and protected it. This basic formulation has served as the foundation of the contemporary science.

We can trace this theme of immunology’s history as a result of those discoveries leading to the elucidation of the bacterial etiology of infectious diseases. This orthodox narrative draws together twin disciplines: microbiology, the study of the offenders, and immunology, the examination of host defense. Thus typical accounts of immunology’s development begin with relating how, on the one hand, German immunochemists—led by Robert Koch’s epochal bacteriological studies—discerned specific antibody protection of infected animals, and on the other hand, Elie Metchnikoff, at the Pasteur Institute, championed cellular immune mediators for similar purpose. These apparently disparate versions of immunity eventually merged, and a unified theory of host defense was outlined. In this view,

both humoral and cellular agents were used for one purpose, to kill pathogens, and the argument seemed confined to questions of mechanism. Other autoimmune behaviors were marginalized as aberrancies, and the evolutionary origins of the immune system were postulated only in regards to the ostensible defensive functions discerned in vertebrate animals. So by 1908, when Metchnikoff and Paul Ehrlich, a stalwart immunochemist, shared the Nobel Prize, the discipline had quickly reached an apparent theoretical maturity amid great clinical fanfare for its promise to combat infectious diseases.

Thus, in this pathological context, immunology began as the study of how a host animal reacts to pathogenic injury and defends itself against the deleterious effects from such microbial insult. This is the historical account of immunology as a clinical science, a tool of medicine, and as such it focused almost exclusively on the role of immunity as a defender of the infected. And note, the paradigmatic host is the patient, an infected “self” which I believe is a critical element for the power of this view. To erect such an edifice is to explain the development of the field by assuming a major presupposition, namely that there is a definable self that might be defended. After all, are we not selves? I have contested that view at its foundations. For me the history of immunology is precisely the very attempt to define such an entity. And the reading I have offered contends that it is the problem of a “self” that has besieged the theoretical fortress immunology has attempted to erect. Briefly put, the issue for me has been neatly divided between those who adopt a notion of the self, namely that there indeed is a “self,” and the contesting argument that the self is an artifice, a conceit, a model at best. In this latter view, immunology cannot accept organismal identity as a given, but must pose the “self” as its core problematic.

The clinical orientation, which assumes a given entity—the self—is obviously a dominant organizing perspective, but it has been challenged by a formulation that works by an altogether different theoretical construction, dispensing with the self altogether. In such views, immunology is not at heart the science that seeks to discern the basis of self/nonself discrimination as current dogma would have us believe, but rather it is concerned with the *establishment* of organismal identity. This issue has been present since the birth of the field, but effectively subordinated to the history of immunological specificity. No doubt this latter problem was a dominant concern, but it rested upon an unstable theoretical construction. The utility of the “self” resides in its own semantic instability, its diverse meanings and uses. But as the science of immunology has evolved, the aspiration to more precisely understand and articulate the science of antibodies and lymphocytes increasingly makes such metaphorical approximations obsolete.

Nineteenth-Century Origins of Immunology

If we begin our inquiry with a careful examination of the earliest immunologists (Elie Metchnikoff, the German microbiologists who debated him, and the immunochemists who built on his formulation), we may clearly discern the origins of the immune selfhood problem [7, 8]. Metchnikoff came to the nascent field of immunology from an unexpected theoretical and methodological perspective. He was an embryologist who sought to discover genealogical relationships in the context of Darwinian problematics. Intrigued with the problem of how divergent cell lineages were integrated into a coherent, functioning organism, Metchnikoff was thus preoccupied with the problems of development as process, which he regarded as Darwinian: cell lineages were inherently in conflict to establish their own hegemony, and he thus hypothesized that a police system was required to impose order, or what he called “harmony” on the disharmonious elements of the animal [7, 8]. He found such an agent in the phagocyte, which retained its ancient phylogenetic eating function, to devour effete, dead, or injured cells that violated the phagocyte’s sense of identity.

Thus the phagocyte was initially viewed as a purveyor of identity, and when Metchnikoff became engaged in the fledgling field of infectious diseases at the beginning of the 1880s (with almost 20 years of comparative embryology to his credit), he was poised to apply his understanding of phagocytes to the duty of protecting the organism from pathogens (i.e., maintaining integrity [2, pp.20, 62–63]). It was a grand scheme, which he presented in a series of public lectures in Paris in the spring of 1891. Later published as *Lectures in the Comparative Pathology of Inflammation*, Metchnikoff argued that the phagocyte had preserved its most ancient phylogenetic function, which in simple animals was to serve as the nutritive organ (eating resident microbes). In animals with a gut, the phagocytes continued to eat, but now for defense. Thus in Metchnikoff’s theory, immunity was a particular case of physiological inflammation, a normal process of animal economy. But there was a more subtle message: (1) immunity was an active process with the phagocyte’s response seemingly mounted with a sense of independent arbitration; and (2) organismal identity was a problem bequeathed from a Darwinian perspective that placed all life in an evolutionary context. This last point cannot be overemphasized: Darwinian conflict extended to the competing cell lineages of the individual animal. The agency quality of his argument and the radical sense of self-definition reflected major Nietzschean themes, a parallel I have attempted to make explicit elsewhere [2, 9, 10].

If we consider the documentation of Metchnikoff’s Nobel archives and public testimony, and carefully contrast his views with other scientists who were involved in similar research, we must recognize that his scientific posture employed emergent and dynamical thinking appropriate to an organ-

ismic orientation of a biologist keenly aware of the problem of identity in a post-Darwinian age [2, 7, 8, 11]. Metchnikoff, perhaps with unique insight, deeply comprehended the Darwinian revolution. He maintained that throughout the life experience of the organism there are changing environments, new insults, and encounters with novel challenges, and it is the immune adaptability and versatility that determines overall success. His concerns were different from those who approached immunology from the more narrow perspective of infectious diseases; Metchnikoff was concerned with the grand challenge of Darwinism writ large. He sought the primary lesson of evolutionary biology, and his was a radically different conception of the organism from that offered in the pre-Darwinian era.

Prior to Darwin's *On the Origin of Species*, the organism was a "given." It was viewed as essentially unchanging and stable, in contrast to the dynamic image implicit in Metchnikoff's formulation. For him the organism is in a dialectical relationship with its world. In an ever-changing set of relationships, at many different levels of engagement, the animal lives in response to its environment, and it was this component of *active* reaction that was so revolutionary. But, because of the apparent independent volition (or agency) of the phagocyte that could hardly be described by physico-chemical laws, he drew the ire of those committed to establishing the chemical basis of the immune reaction. For that posture, he was trampled underfoot by the growing tide of immunochemistry. He challenged the simpler formulation that there was an entity that could be defined; in a sense, this was a necessary pre-condition for those who would seek to demonstrate those reactions that conferred protection of such an entity. If the entity itself was in question, the entire enterprise assumed a complexity not readily unravelled. For the time, it was sufficient to make the assumption that indeed there was a "self," and then proceed. Be that as it may, the question of identity remained at the heart of his theory, and more to the point, Metchnikoff's hypothesis was prescient. As an embryologist, he regarded the organism as fundamentally in a continuous developmental mode, and a dialectical approach is required to understand such development. Without firm boundaries and structure, this vision of organismal identity was decentered and emergent [7, 12]. Twentieth-century theoretical immunology rested upon this ambiguity.

The Problematic Immune Self

This dialectical mode of thinking reappeared in immunology almost a century after Metchnikoff's first proposal: Francisco Varela, Antonio Coutinho, and their colleagues (the Paris School) evocatively espoused a self-determinism closely related to Metchnikoff's dialectical vision of the organism: "The self is not just a static border in the shape space, delineating

friend from foe. Moreover, the self is not a genetic constant. It bears the genetic make-up of the individual and of its past history, while shaping itself along an unforeseen path” [13, p. 363]. This dynamic vision of immune identity had a complex, and revealing, history, one that was developed in reaction to the theory of Frank Macfarlane Burnet.

Burnet was a virologist by original training, and he thus came to immunology from a perspective quite different from that of the immunologists then dominating the field. He was ambitious to integrate developmental biology, genetics, and immunology into a cohesive theoretical whole, and he did so by drawing upon both Metchnikoff and later ecological theory to devise a view of the immune system as the purveyor of organismal identity [2, 14]. But if this history is regarded through a prism similar in design to that used to dissect and reconstruct the Metchnikovian saga, namely an ethos that relied on appreciating the dynamical and hierarchal properties of biological systems, one must conclude that Burnet was not the best champion to carry the Russian’s mantle, for Burnet, assuming the Ehrlich precept, sought a firm definition of the immune self. Indeed, given the same orientation as Metchnikoff’s major early competitors, the microbiologists, we can readily appreciate that Burnet’s own training and research interests fell squarely in their tradition. So while Burnet’s concerns were dominantly biological and not chemical, he still regarded the host organism as an entity, and immunity was to defend it.

Burnet’s theory, in brief, was that during prenatal development, the animal exercised a purging function of self-reactive lymphocytes (the cells responsible for synthesizing reactive antibodies and mediating so-called cellular reactions), so that all antigens (substances that initiate immune responses) that were encountered during this period would be ignored by the immune system. The hypothesis, first presented in 1949, was later developed into the clonal selection theory, which maintains that lymphocytes with reactivity against host constituents are destroyed during development, and only those lymphocytes that are non reactive would be left to engage the antigens of the foreign universe [15, 16]. These potentially deleterious substances would select lymphocytes with a high affinity for them, and through clonal amplification a population of lymphocytes would differentiate and expand to combat the offending agents. The theory was assumed to be proven in the 1970s, but this vision of immune identity should be regarded with considerable skepticism.

Bountiful evidence in recent years has suggested that autoimmunity is a normal finding, and that immune reactivity, rather than functioning only in an “other-directed” mode, is in fact bidirectional:

Clearly, one can define “self” from a biochemical or genetic or even a priori basis. But from our vantage point, the only valid sense of immunological self is the one defined by the dynamics of the network itself. What does not enter into its cognitive domain is ignored (i.e., it is non-sense). This is in clear contrast to the traditional

notion that IS [immune system] sets a boundary between self in contradistinction to a supposed non-self. From our perspective, there is only self and its slight variations. [13, p.365]

The critical turning point was to appreciate that the immune system in fact recognizes selfness as natural autoimmunity, and such host-directed reactivity is physiologically normal. The significance of this orientation has taken some time to sink into the collective consciousness of the discipline, and its ramifications are still not widely appreciated. As Coutinho and Michel Kazatchkine later wrote:

During this century, the evolution of concepts on autoimmunity could be summarized by “never, sometimes, always.” Thus from the early “horror autotoxicus” [Ehrlich] to the 1960s, immune autoreactivity was simply not considered. . . . With the first identification of autoreactive antibodies in patients and the subsequent conceptual association with autoaggressive immune behaviors, the “sometimes” phase was entered, necessarily equated with disease. By this time, immunology had laid its foundation on the clonal selection theory, which forbids autoreactive clones in normal individuals. Immunologists thereafter devoted 30 years discovering ways by which autoreactive lymphocyte clones can be deleted and why they fail to be deleted in autoimmune patients. . . . In the 1970s at least three sets of observations and ideas began to alter this course of events and to herald the “always” period. [17, pp.1–2]

This position contrasts with the “one-way” definition of selfhood, where there is a genetic self whose constitutive agents see the foreign, and immune reactivity arises from this polarization with attack directed only against nonself. There are at least half a dozen different conceptions of what constitutes the immune self, versions that might be situated by where one places the model on a continuum between a severe genetic reductionism and a complex construct employing different principles of organization [18]. With so much dispute surrounding the definition of self, a growing counterposition suggests that the “self” might be better regarded as only a metaphor for the immune system’s silence, i.e., its non-reactivity. The theory built upon the “self” now appears to have many ad hoc caveats and paradoxes. Perhaps the evolution of the original metaphor to theory is now yielding again to another metaphorical construction. So what could the Immune Self be? A radical answer was offered by Niels Jerne.

Jerne and the Deconstruction of the Immune Self

Jerne went well beyond the current notion of the immune network composed of lymphocyte subsets, secreting immuno-stimulatory and inhibitory substances (essentially a simple mechanical model with interlaced, first-order feedback loops), to propose a novel conception of immune regulation. His idiotypic network theory was, from its very inception, a complex amalgam of fitting the pieces of the regulation puzzle in place, with an

overriding desire of understanding the immune system as a cognitive enterprise. The idiotypic network theory hypothesis, extensively presented in 1974 and proposed in outline earlier, proposed that antibodies formed a highly complex interwoven system, where the various specificities “referred” to each other [19, 20]. Under the general rubric of “cognition,” Jerne conceived of the immune system as self-regulating, where antibody not only recognizes foreign antigen, but is capable of recognizing self constituents as antigens (the so-called idiotopes). There was no essential difference between the “recognized” and the “recognizer,” since any given antibody might serve either, or both, functions. In other words, immune regulation was based on the reactivity of antibody (and later lymphocytes) with its own repertoire, forming a set of self-reactive, self-reflective, self-defining immune activities. There is no “self” and “other” for the immune system; according to Jerne’s theory the system is complete unto itself, consisting of interlocking recognizing units: each component reacts with certain other constituents to form a complex network or lattice structure. When the system is perturbed by the introduction of a substance that is “recognized” (i.e., it reacts with a member of the system), this disturbance initiates immune responsiveness. Thus foreignness per se does not exist in this formulation.

Jerne’s theory presents a radically altered view of immune selfhood. If the biological world was so easily divided between “host” and “foreign” constituents, then anything an antibody (or lymphocyte) encountered would be suitable for destruction according to Burnet’s original tolerance theory. In that simplified world of self/nonself discrimination, the immune system learned such distinctions, generated an army of reactive antibody and lymphocytes, and acted accordingly when “antigen” was encountered. But Jerne coupled the simple antibody-antigen interactions with the far more complex and non-discriminatory functions of the immune system that built upon self-recognition. Thus “autoimmunity” became the organizational rule to explain immune function. The idiotypic network was fundamentally “dynamic” and “self-centered,” generating anti-idiotypic (“self-reactive”) antibodies to its own antibodies, which he thought constituted the overwhelming majority of antigen present in the body [21, 22]. Strikingly, there is no explicit mechanism for self/nonself discrimination, and this apparent lacuna served as the nexus of critiques (e.g., the arguments of Melvin Cohn [23, 24]). But for Jerne, the need to define the self as distinct from the other receded from his primary theoretical concerns, and this posture was to have important repercussions.

When the immune system is regarded as essentially self-reactive and interconnected, the “meaning” of immunogenicity—that is, reactivity—must be sought in some larger framework. Antigenicity then is only a question of degree, where “self” evokes one kind of response, and the “foreign” another, based not on its intrinsic foreignness, but rather because

the immune system sees that foreign antigen in the context of invasion or degeneracy. There is no foreignness per se, because if a substance was truly foreign, it would not be recognized—there would be no image by which the immune system might engage it. So the “foreign” becomes a perturbation of the system; as observers, we record the ensuing reaction, and only as third parties do we designate “self” and “nonself.” From the immune system’s perspective, it only knows itself, and thus reaction to the foreign becomes secondary, or perhaps a by-product of this central self-defining function. The hypothesis served as a fecund nexus for the reformulation of the entire question of how the immune system is organized. If there is a “self” in Jerne’s theory, it is the entire immune system as it “senses” itself. Jerne’s theory thus appears radically different from the dominant theories of interlocking inhibitory/stimulatory activities that described immune function built from Burnet’s original self/nonself dichotomy. Jerne’s theory commanded a rethinking of the entire conception of self/nonself.

The challenge posed by Jerne’s network theory was two-fold: most generally, it demanded an assessment of an inward-directed self-seeking process; its critical weakness was lacking a stable reference for defining selfness. Later models arising in reaction to his network theory attempted to circumvent this latter problem by demoting the problem of selfhood altogether. It is in this fundamental reorientation—if one takes it seriously—that one may perceive a decisive shift in immunology’s theoretical foundations.

One does not require Jerne’s particular theory of the immune network to see that the more basic principle of the context of antigenicity is the critical governing parameter of immune reactivity. Consider the dominant model concerning lymphocyte activation, where it is generally appreciated that specific recognition of antigen by a lymphocyte receptor is not sufficient for activation, and that additional signals determine whether a cellular response or cell inactivation follows. In short, an antigen is neither self nor nonself except as it attains its “meaning” within a broader construct [25, 26]. Orthodox immune theory encompasses this idea in the so-called “two-signal model,” which does not require any of Jerne’s hypotheses to fulfill its agenda [23, 24]. But there are more radical readings of the “contextualist” setting by which antigens are sensed. Debate concerning what constitutes the milieu of “meaning” of antigenicity and ensuing reaction have spawned certain provocative, and potentially important, models of immune regulation. I believe that the novel responses inspired by Jerne’s theory herald a shift in the very foundations of how immune regulation might be understood.

The entire contra-Burnetian perspective rests on recognizing the “relativity” of perspectives, for the context of the immune encounter is paramount in conferring meaning on any antigen. As Yves Borel commented at an important tolerance workshop in 1986:

[T]oo much emphasis has been put on the self vs. nonself discrimination. From the point of view of the antigen this is a misnomer; the antigen is neither self nor nonself and I think that, as Landsteiner has elegantly demonstrated, the antigen could be a hapten [a substance that becomes antigenic only when coupled to a larger molecule, e.g., a protein]. A hapten is neither self nor nonself. Depending on the readout system in which the determinant is seen, it becomes self or nonself. So self/nonself is not defined from the point of view of the antigen but from the point of view of the host and it is really the context of recognition which defines self and nonself. [27, pp.23–24]

Once the self/nonself structure is weakened, more iconoclastic perspectives may be entertained. For instance, Zvi Grossman has proposed that under the cover of immunological tolerance autoreactive T lymphocytes may actively participate in classical protective responses, as well as in non-classical physiological activities, such as stabilizing the differentiation profiles of other cells [28]. He has argued:

[t]he immune system is not “devised for aggression against foreign antigens” more than it is devised to manifest tolerance, or [a more] complex relationship, to self or foreign antigens; recognition of antigen is necessary for both aggression and tolerance but is not sufficient for either. [29, p.47]

This credo is developed most explicitly by Polly Matzinger, who arrives at her model from a reaction to the lymphocyte activation hypothesis rather than as a development of post-Jernian theorizing [18]. In her view, the immune system decides what is insulting to the organism—i.e., what causes distress, destruction, or non-programmatic death—and through signals of such aberrancy, immune reactions are initiated. Selfhood, per se, recedes as the basis of immune definition. Immunity becomes organismally driven (i.e., functional) and immunocytes become dependent on extra-immune factors and context.

In Matzinger’s theory, self/nonself differentiation has thus been replaced with a contextualist scheme based on responses to danger and destruction. From this perspective, both the self and its immune system have been deconstructed. Rather than a specialized system that patrols the rest of the body, the immune system becomes extended and intricately connected with every other cell of the organism so that tolerance and reactivity are governed by the cooperation of lymphocytes, antigen-presenting cells, and other tissues. Matzinger’s model is fundamentally a process-driven, functionally conceived model, and interestingly, it builds on the antigen-presenting cell (likely a close relative of Metchnikoff’s original phagocyte) as the arbiter of immune reaction, a cell that cannot distinguish self from nonself in traditional terms of lymphocyte recognition. Thus to varying degrees some contemporary immunologists are moving beyond the self as a governing principle, and replacing self/nonself discrimination with a functional, process-dominated conception based on the context in which antigen is encountered.

The differences in the way the cognitive or contextual orientation takes shape can be linked to the different elements with which the various theorists develop the immune–nervous system analogy, each assuming a distinctive slant from Jerne’s own formulation. Irun Cohen has singled out the imprinted knowledge of the external world, a knowledge stored developmentally in sets of molecules and their complementarities, and, to a limited extent, in simple regulatory networks [30]. Coutinho and Varela have focused on the distributed nature of information processing and self definition; Grossman has emphasized “learning from experience,” arguing that the coordination of complex physiological activities of many cells, across tissues and several levels of organization, cannot be strictly pre-programmed, nor restricted to structures exclusively confined to the immune system. Matzinger dispenses with the self concept altogether, focusing instead on “danger” and “damage control” [13, 28, 29, 18]. In their scrutiny as to how well the “one-way”—i.e., self against the other—paradigm might account for selfhood, these critics decided that the entire theory required radical revision [31]. The self no longer commands center stage.

Why Does the Immune Self Linger?

So now, with this background, let us return to Howes’ critique of my reading of this history and the state of contemporary immunology [6]. As already mentioned, the question of immune identity was only an implicit issue at the birth of immunology in the late 19th century, but it indeed served as the nexus of Metchnikoff’s phagocytosis theory, which explicitly posed the problem of organismal identity as the core problem of immune reactivity. His was an elaborate theory born in the throes of early Darwinism, and the concept of the immune self, although not articulated as such, grew from his previous research in developmental biology (summarized here in 1991) [7, 8]. His history is important for understanding the “infrastructure” of immunology from its earliest inception; and, although his ideas were quickly eclipsed by the development of immunochemistry and serology, Metchnikoff’s organismal orientation was to resurface after World War II, when autoimmunity and transplantation became crucial clinical concerns [1]. This is the context within which Burnet first used the term “self” in 1949, as an explicit vehicle for building a theory of immune reactivity [2, 14, 15]. *The Immune Self, Theory or Metaphor?*—the focus of Howes’ article—is a study of theory construction and language, specifically how “self” with its various meanings was co-opted from philosophy and psychology in order to serve immunologists with a ready concept upon which to build a scientific theory [2]. The term was not widely used until the 1970s; by that time, not coincidentally, Burnet’s clonal selection theory had be-

come dogma [25]. The notion of an immune self, that which identified the targets of immune reactivity, became the defining matrix of the discipline, then dubbed the science of self/nonself discrimination.

In my examination of the concept of the self, I attempted to show how immunologists used “self” in its different modalities for their theorizing. Whether consciously or not, immunologists extrapolated from their common experience (formalized in philosophical discussion from Descartes to the postmodernists) various uses of the term that they then used to structure the science. In this sense, “self” has been highly plastic, serving various rhetorical and conceptual functions. I then analyzed and critiqued the use of the self as a concept in contemporary immunology (i.e., post-1974 [Jerneian network theory]), offering both philosophical arguments and more technical analyses [2, 3, 25, 32–35]. I cannot further dwell on those detailed discussions here. Instead, I will now address the main concern of Howes’ critique of my position.

The issue, and the one Howes correctly focuses upon, is whether self might be defined. Whether considered as a philosophical construct or a medico-scientific one designed for immunology, its cognitive meaning rests on its philosophical interpretation. On this Howes and I agree. I maintain that “self” is hopelessly obscure, and she concurs when she admits that “philosophers have utterly failed to define the self” [6, p.120]. But, if Howes concedes that we cannot define the self, then what are we discussing? That is precisely the philosophical argument, and I will turn to it. But first let us note that a failure to define “self” hardly signifies that the word lacks important uses. Rather it only signifies that its indistinct cognitive status—its “fuzzy” or “elusive” borders—are, in fact, an important part of its utility—indeed, its power. But she is dissatisfied with that answer, and moves from my acceptance of the self’s vague metaphorical status to seemingly wishing to salvage a more definite standing for it. Howes seems to want to reclaim something: “Many of us, however, are unprepared to accept that there is no self at all” [6, p.120]. Of course, I never said that there are no “selves,” for obviously we *are*, but my immediate question concerned the term’s meaning and standing in immune theory.

Howes, building on my presentation of Niels Jerne’s network theory and its development, correctly reads me as suggesting that a process-oriented approach to understanding the immune system is fecund and promising, yet she misses the key point: Jerne dispensed with the self/nonself distinction altogether, (although some of his followers, e.g., the so-called Paris School, perhaps inadvertently allowed it to slip it back in) [3, 7, 11]. From Jerne’s perspective, the self/nonself dichotomy slips away, and definitions of the immune system are then given as functional assessments—reactive or quiescent—a point of view now supported by increasing scientific evidence [25, 34, 35]. Why is the self still important to Howes? Granting the utility of the term during a period in the evolution of immunology’s history,

we have good reason to question its continued centrality as new challenges to Burnet's self theory appear. Given our inability to structure a theory that defines "self" and "foreign" in contemporary immunology, many scientists have satisfied themselves with functional responses, that is, simply understanding the immune system's "on" and "off" behavior in molecular terms. When writing *The Immune Self*, I predicted this development, initially not because of the science (there was little concrete evidence in 1991–92), but on the basis of the inability of philosophy to define the self as an *entity*. I freely drew on postmodern criticism, which only brought a much older critical tradition to the fore [32]. Beyond examining this contemporary assessment of identity functions in immune theory, we do well to explore the history of the concept of self. In *The Immune Self* and again here, I will borrow Charles Taylor's description of the "punctual self" to make the point [36].

John Locke attempted to construct the "self" as an autonomous legal unit to fulfill certain 17th-century liberal political goals. To do so he extrapolated from a philosophical invention. The "punctual self" was part of the early modern scientific conceit that regarded the knowing subject as totally divorced from the world in order to objectify it. The self also became a subject of scrutiny, so that ordinary experience might be seen from afar, objectified, and thereby controlled. Locke thereby reified the mind to an extraordinary degree, adopting a kind of atomism. This construction of the thinking subject supported a radical disengagement of oneself with a view toward remaking the world: "the real self is 'extensionless'; it is nowhere but in the power to fix things as objects" [36]. This power reposes in consciousness, a theme we can trace to Hume and later to William James and the phenomenologists [2].

Taylor argues that the self, as an object (or entity) must fulfill certain criteria: it must have objective status, standing independent of any description or interpretation, and be capturable in explicit description. Most importantly, an object can and must be understood without reference to its surroundings or contingent circumstances [2, pp.34–35]. The self of Locke and David Hume (whom Howes specifically invokes) cannot fulfill these criteria. Taylor makes the critical point that the "self" is the answer to the question "Who am I?" which requires answers totally dependent on cultural or moral contexts, frameworks, or orientation—human categories of personal and social action, of value. Epistemological criteria are obviously operative, but the question of personal identity is a *moral* issue, not an epistemological one. The self, then, is a moral description or category, one that fulfills criteria of human identification.

Thus the "punctual" or "detached" self arose from two tributaries: the agent of a moral philosophy in post-Reformation England and the scientific actor who would scrutinize the world, and himself, apart from any constitutive concerns: The self's "only constitutive property is self-awareness. This

is the self Hume set out to find and, predictably, failed to find” [36, p.50]. This identity over time, some psychological continuity that Howes would invoke as “an imperfect identity,” is a construction that simply fails to fulfill the essential criteria of an entity given above. Most importantly, defining personal identity, the “Who am I?” question, can only be answered in the context of social and moral concerns. Context-determined, the self simply cannot be reduced to some single psychological construction of continuity, not only because self-consciousness cannot be captured or defined, but more fundamentally, as Hume himself realized, because this was only a psychological conceit, a vague awareness of an identity function held together by psychological reflection. It was a useful invention, but it had no basis that could be defined by any analytical criteria.

So let us return to Howes’ thesis. She wants to salvage the self, and besides declaring that there are selves, she remarkably presents no argument in support of her wish. In a move that I find baffling, she simply dissociates the “self” from “personal identity” and incorrectly attributes this “solution” to Hume: “Hume’s solution to the problem of personal identity is to make personal identity irrelevant to having a self” [6, p.121]. There are two issues to untangle. First, while correctly reading Hume’s notion of personal identity as a psychological construction in which we “see” continuity over time, she apparently fails to understand Hume’s view of the self as “imperfect” as anything less than an admission of its elusive metaphysical character. Philosophically, “imperfect” is a kind way of saying “ungraspable” (by the criteria listed above)—i.e., the “self” does not exist in any analytic sense. This was hardly the “punctual” self Locke strove to capture. But then the second matter, perhaps even more fundamental, is, if Howes dissociates personal identity from the self (I believe a wrong inference from Hume), then what *is* the self? She can only answer, “We should abandon the notion of perfect personal identity, then, and retain belief in the self” [6, p.122]. An “imperfect personal identity”? Because it is “imperfect,” it is undefinable: This is perhaps acceptable in certain discourses, but science strives for something else.

I think we all know what Howes means when we refer to the “self” as a human agent, one immersed in human activities and concerns. “Self” is a moral category we use to describe a social and personal identity. But the point of the discussion is not the self’s human or moral standing, but rather its use as a scientific construct in immunology today, which I maintain is problematical. When we use such language, we cannot point to any such entity and define it, which is why I called it “elusive” [2]. To be sure, there has been metaphorical utility in the term, but in science we seek to move from metaphor to more concrete and precise definitions. If we recognize that our theory is based on phenomena that may indeed be described more objectively, then we adopt those terms of discussion, because we seek a scientific lexicon that more precisely refers to natural entities and their

relationships. And if our metaphors become distorting to the evidence, we abandon them.

We are “selves,” but the immune system is a biological activity. And most saliently, the immune system is not a *human category*. We make a category error in assigning human descriptions to lymphocytes and antibodies, which reside in their own domain, objectified, possessing no self-consciousness as we understand our own psyches, and hopefully freed from our possessive prejudices. Immunology is now moving beyond its old metaphors, embracing new theories and new metaphors (“danger” [18]) to build its theoretical edifice. Perhaps in its failure—showing us that the “self” cannot be defined epistemologically—immunology has forced us to more carefully assess what we mean by selfhood and thereby delve more deliberately into its true character—its moral meanings.

REFERENCES

1. SILVERSTEIN, A. *A History of Immunology*. San Diego: Academic P, 1989.
2. TAUBER, A. I. *The Immune Self: Theory or Metaphor?* New York: Cambridge UP, 1994.
3. TAUBER, A. I. The molecularization of immunology. In *The Philosophy and History of Molecular Biology: New Perspectives*, edited by S. SARKAR. Dordrecht: Kluwer Academic Publishers, 1996. 125–69.
4. GOLUB, E. S., and GREEN, D. R. *Immunology: a Synthesis*, 2nd ed. Sunderland, MA: Sinauer, 1991.
5. KLEIN, J. *Immunology*. Boston: Blackwell, 1990.
6. HOWES, M. The self of philosophy and the self of immunology. *Persp. Biol. Med.* 42:118–30, 1998.
7. TAUBER A. I., and CHERNYAK L. *Metchnikoff and the Origins of Immunology: From Metaphor to Theory*. New York: Oxford UP, 1991.
8. TAUBER, A. I. The immunological self: A centenary perspective. *Persp. Biol. Med.* 35:74–86, 1991.
9. TAUBER, A. I. The organismal self: Its philosophical context. In *Selves, People, and Persons*. Boston University Studies in Philosophy and Religion. Vol. 13. Edited by L. ROUNER. Notre Dame: Notre Dame UP, 1992. 149–67.
10. TAUBER, A. I. A typology of Nietzsche’s biology. *Biol. Philos.* 1994. 9:25–44, 1994.
11. TAUBER, A. I. The birth of immunology. III. The fate of the phagocytosis theory. *Cell. Immunol.* 139:505–30, 1992.
12. CHERNYAK, L., and TAUBER, A. I. The idea of immunity; Metchnikoff’s metaphysics and science. *J. Hist. Biol.* 23:187–249, 1990.
13. VARELA, F. J.; COUTINHO, A.; DUPIRE, B.; and VAZ, N. N. Cognitive networks: Immune, neural, and otherwise. In *Theoretical Immunology, Part Two*, edited by A. S. PERELSON. Redwood City: Addison-Wesley., 1988. 359–75.
14. TAUBER A. I., and PODOLSKY S. H. Macfarlane Burnet and the declaration of the immune self. *J. Hist. Biol.* 27:531–73, 1994.
15. BURNET, F. M., and FENNER, F. *The Production of Antibodies*. 2nd ed. Melbourne: Macmillan, 1949.
16. BURNET, F. M. *The Clonal Selection Theory of Acquired Immunity*. Nashville: Vanderbilt UP, 1959.

17. COUTINHO, A., and KAZATCHKINE, M. Autoimmunity today. In *Autoimmunity: Physiology and Disease*, edited by A. COUTINHO and M. KAZATCHKINE. New York: Willey Liss, 1994. 3–6.
18. MATZINGER, P. Tolerance, danger, and the extended family. *Ann. Rev. Immunol.* 12:991–1045, 1994.
19. JERNE, N. K. The immune system. *Scientific American* (July): 52–60, 1973.
20. JERNE, N. K. Towards a network theory of the immune system. *Ann. Inst. Pasteur/Immunology* 125C:373–89, 1974.
21. JERNE, N. K. The generative grammar of the immune system. *EMBO J.* 4: 847–52, 1985.
22. JERNE, N. K. The generative grammar of the immune system. In *Prix Nobel 1984*. Stockholm: Almqvist and Wiksell, 1985.
23. COHN, M. Conversations with Niels Kaj Jerne on immune regulation: Associate versus network recognition. *Cell. Immunol.* 61:425–36, 1981.
24. COHN, M. The concept of functional idiotype network for immune regulation mocks all and comforts none. *Ann. Inst. Pasteur/Immunology* 137C:64–76, 1986.
25. PODOLSKY, S. H., and TAUBER, A. I. *The Generation of Diversity: Clonal Selection Theory and the Rise of Molecular Immunology*. Cambridge: Harvard UP, 1997.
26. TAUBER, A. I. Historical and philosophical perspectives on immune cognition. *J. Hist. Biol.* 30:419–40, 1997.
27. BOREL, Y. Comments to Melvin Cohn: The ground rules determining any solution to the problem of the self/nonself discrimination. In *The Tolerance Workshop: Proceedings of the EMBO Workshop on Tolerance*, edited by P. MATZINGER, H. G. RAMENSEE, G. STOCKINGER, et al. Basel: Editiones (Roche), 1986.
28. GROSSMAN Z., and PAUL, W. E. Adaptive cellular interactions in the immune system: The tunable activation threshold and the significance of sub-threshold responses. *Proceed. Nat. Acad. Sci. (USA)* 89:10365–69, 1992.
29. GROSSMAN, Z. Cellular tolerance as a dynamic state of the adaptable lymphocyte. *Immunol. Rev.* 133:45–73, 1993.
30. COHEN, I. R. Kadishman’s tree, Escher’s angels, and the immunological homunculus. In *Autoimmunity: Physiology and Disease*, edited by A. COUTINHO and M. D. KAZATCHKINE. New York: Wiley-Liss, 1994. 7–18.
31. STARZL, T. E., and DEMETRIS, A. J. Transplantation milestones viewed with one- and two-way paradigms of tolerance. *JAMA* 273:876–79, 1995.
32. TAUBER, A. I. Postmodernism and the immune self. *Science in Context* 8: 579–607, 1995.
33. CHERNYAK L., and TAUBER A. I. The dialectical Self: Immunology’s contribution. In *Organism and the Origins of Self*, edited by A. I. TAUBER. Dordrecht: Kluwer Academic Publishers, 1991. 109–56.
34. TAUBER, A. I. Conceptual shifts in immunology: Comments on the “two-way paradigm.” In *Paradigm Changes in Organ Transplantation*, edited by K. F. SCHAFFNER and T. E. STARZL. *Theor. Med. Bioethics* 19:457–73, 1998.
35. TAUBER, A. I. Response to Melvin Cohn: How Cohn’s two-signal model was turned. In *Paradigm Changes in Organ Transplantation*, edited by K. F. SCHAFFNER and T. E. STARZL. *Theor. Med. Bioethics* 19:485–94, 1998.
36. TAYLOR, C. *Sources of the Self: The Making of the Modern Identity*. Cambridge: Harvard UP, 1989.