PART IX

AFTERWORD

Wayne D. Gray

Someone has to go last and in this case the honor falls to Mike Byrne. I did not commission Mike to write “a last chapter” but that is what he did. In a way his chapter 30 and my chapter 1 can be viewed as bookends for the 28 chapters that lie between. I began the book with a short essay on three types of control of cognition that need to be considered when building integrated models of cognitive systems. Not wanting to overfit the data, I did not try to sharply define what these distinctions were. Instead my goal was to sketch some broad distinctions that might make conversations between research camps easier and, indeed, might give those who build integrated models a new vocabulary to discuss their models and to communicate their needs to those who build single-focus models of cognitive functions or, as Mike puts it, local theories.

Before writing chapter 1, my discussion of Type 1, 2, and 3 control was limited to several paragraphs on the now defunct workshop Web site. I am pleased that, in addition to Mike’s chapter, three other chapters have found a good use for this vocabulary (Gluck, Ball, and Krusmark in chapter 2; Gunzelman, Gluck, Price, Van Dongen, and Dinges in chapter 17; and Ritter, Reifers, Klein, and Schoelles in chapter 18). I am especially pleased that Mike seems to have used the Type 1 versus Type 2 distinction as a call-to-arms for the cognitive community to come together and start building the types of models needed to create integrated models of cognitive systems. His perspective is that of a Type 1 modeler in search of Type 2 components that can be plugged and played in his models. His appeal is very direct, very scholarly, and very personal.
Reasons for considering integrated models of cognitive systems as a desideratum are myriad, ranging from deeply theoretical interest in the structure of cognition for its own sake to the desire to gain leverage on difficult applied problems to pedagogical interests in blending cognitive psychology with related disciplines like artificial intelligence. I would like to concentrate on the more applied reasons, many of which are also articulated elsewhere (see Byrne, 2003; Byrne & Gray, 2003).

Consider a modern jetliner pilot. The task faced by the pilot is complex, safety critical, and takes many years of intensive training to master. The environment in which this task is done is staggeringly complex, visually rich, places extreme time demands on the flight crew, and is partially managed by a fairly opaque piece of automation. This is not only a challenging domain for the pilots, but for human factors engineers who have the task of trying to make the pilot’s job easier and safer.

As much of this volume might suggest, it is the integration of all these capabilities which is particularly pertinent here. Scientists and engineers who want to understand pilots, or emergency room physicians, or even just regular people doing fairly routine office work (e.g., surfing the Web) do not have the luxury of isolating one aspect of human performance. Real people in real settings—to borrow Ed Hutchins’s (1995) terrific turn of phrase, those exercising “cognition in the wild”—bring to bear an integrated set of capabilities. Within the space of a minute, a pilot or an office worker might retrieve something from memory, plan, conduct a visual search, execute a routine procedure, make a linguistic inference, produce a judgment, and solve a problem. These activities are not circumscribed but rather highly dependent on one another. A key inference in resolving a language comprehension ambiguity might be driven by the outcome of a visual search and the new information comprehended lead to a new plan of action. People who study pilots, doctors, and office workers cannot specialize in one small area of psychology but rather must confront human cognition, perception, and action as an integrated unit. Furthermore, this serves as a reminder: the human mind, the...
thing we are trying to understand and model, actually functions as an integrated whole.

Oddly, experimental psychology has tended not to view the problem this way. The dominant approach has been “divide and conquer.” This may serve well in some contexts, but there are potentially serious hazards. Even though it is true that human visual attention can operate in impoverished visual contexts for trivially simple tasks, it is not necessarily the case that visual attention operates the same way in those circumstances as it does when the person is simultaneously trying to monitor multiple flight instruments and make a decision about whether to abort a landing. Nonetheless, experimental participants in psychology laboratories have done literally millions of trials where the entire experience consists of making a present/absent judgment on red or green horizontal or vertical bars on uniform black or white backgrounds. I would be unsurprised to find out that the number of paired associates memorized in psychology laboratories numbered in the billions. I am not trying to suggest that such research has no value; I think we have learned a great deal about visual attention from search experiments and have learned useful things about human memory from list memorization experiments. No, the problem is that accounts based on experiments like these tend to stop at the boundaries of those experiments. They do not fit in with other theories and accounts; there is no coherent picture being formed.

This is precisely the problem Allen Newell (1973) warned us about more than 30 years ago in his famous “twenty questions” paper. This paper should be required reading for anyone studying human cognition because of prophetic comments like this one:

I submit that, in the next thirty years we continued as we are now going, another hundred phenomena, give or take a few dozen, will have been discovered and explored. Another forty oppositions will have been posited and their resolutions initiated. Will psychology then have come of age? Will it provide the kind of encompassing of its subject matter—the behavior of man—that we all posit as a characteristic of a mature science? And if so, how will this transformation be accomplished by this succession of phenomena and oppositions? (pp. 287–288)

Suppose that in the next thirty years we continued on for the past 30-odd years more or less as it was in 1973 (though certainly with a variety of improvements, particularly methodological) and that it has largely not made it to the level of mature science. Theories in cognitive psychology are generally too informal and too piecemeal to be considered on the same playing field as physics, chemistry, and even modern biology. This is because the divide-and-conquer approach is still the prevalent mode of operation. Perhaps surprisingly, this is often understandable. The human cognitive/perceptual/motor system is obviously incredibly complex, perhaps the most complex system we humans have ever tried to understand. The problem of understanding it all is so vast; it seems intractable taken all at once. Specialization happens in all sciences by necessity, and the science of cognition is no exception. It is legitimately impossible to be an expert in every aspect of human performance. As I said at the workshop, let the vision folks figure out the gory details of texture segmentation and let the memory folks figure out the gory details of modality-specific short-term memory decay. Divide and conquer has legitimate appeal; I certainly do not want to have to become an expert in coarticulation effects or a myriad of other topics in cognitive psychology.

However, there can be serious drawbacks to the divide-and-conquer approach as an overall research strategy. Divide and conquer is only appropriate to the extent that subsystems are truly independent. There is obviously some validity to this assumption, as there are clearly subsystems that are indeed somewhat independent. For example, it would seem a very poor way to build a motor control system if the machinery that computes the trajectory my arm takes when I reach for the mouse is dependent on whether the sentence I just comprehended was heard versus seen. The issue, then, is how to recover an integrated system from a collection of subdivided parts and to show the value of bothering with the integration.

One integrated approach that is intended to provide insight into numerous interactive domains (and commercial jetliners are most certainly interactive, at least from the perspective of the pilot if not the passengers) is a framework known as the interactive behavior triad, or IBT. The interactive behavior triad is depicted in Figure 30.1 and can be considered a broader variant of a perspective first sketched elsewhere (Byrne, 2001; Gray, 2000; Gray & Altmann, 1999).

The idea behind the IBT is that fully understanding human interactive behavior requires understanding of three crucial sets of constraints. The task is taken to be the goal, or frequently set of goals, the human (or, to adopt human factors jargon, the “operator”)
operator is trying to accomplish and the knowledge required to fulfill those goals. This is not unlike the kind of decomposition found in Card, Moran, and Newell (1983) though perhaps broader in some senses. Misunderstanding the task faced by people is a common source of failure for technological systems. Engineers and computer scientists often invent technologies that, while interesting in their own right, fail to address appropriately any real task actually done by people “in the wild” (more on this shortly). The environment is taken to be the set of constraints and affordances available to the operator from the task environment. This includes not only the various artifacts in the environment but also the properties of any system being controlled directly or indirectly by the operator (e.g., the aircraft flight dynamics, which may set severe limits on how the pilot can attempt to accomplish his/her goals), as well as the broader environment in which the interaction is situated (e.g., general properties of airports; see also chapter 14 by Kirlik and chapter 11 by Todd and Schooler, this volume).

Why are these important to a theory of cognition? A wonderful illustration of a series of fundamental failures to understand both the tasks and environments occupied by real office workers is presented in Sellen and Harper’s (2001) *The Myth of the Paperless Office*. They describe how efforts to achieve the goal of a paperless office environment (explicitly stated by multiple information technology companies over the years) are seriously misguided because such efforts are based on a shallow and inaccurate understanding of both the tasks done by office workers and the important physical and cultural characteristics of offices and the paper in them.

This brings us at last to embodied cognition. This is taken to be the capabilities and limitations of the integrated human perceptual–cognitive–motor system. It is impossible to construct a device (such as a cockpit) and expect smooth interaction with humans without some understanding of how humans themselves work. What is remarkable is how much progress has been made by human factors researchers and practitioners in domains like commercial aviation despite the profound lack of a good integrated theory of perception, cognition, and motor control. In fact there is a tendency for people in such domains to specialize in the task domain rather than in what kind of cognition is done, hence journals and organizations such as “aviation psychology.” Now, it is obviously possible to specialize somewhat within such domains—there are certainly people who focus on specific problems in the visual tasks faced by pilots or radiologists or whatever—but, realistically, complete solutions to problems of interactivity require an integrated view. For reasons laid out in some detail in Byrne (2003) and Byrne and Gray (2003), my strong preference is that this integrated view also meets the standard of being an executable simulation model that therefore makes quantitative predictions.

While one of the aims of the interactive behavior triad is to provide leverage on applied problems, since such problems often hinge on critical interactions between two or more of the IBT components, its use as a guiding framework would also address Newell’s warning. The interactive behavior triad is intended to foster both cumulation and integration because it encourages a “big picture” approach. Consideration of the effects of different tasks and environments discourages the relentless pursuit of repeated minor variations of the same task in the same context. Furthermore, the view of cognition as fully embodied encourages pursuit across long-standing subdiscipline boundaries that are
surprisingly high in experimental psychology; for example, the divide between perception and cognition is amazingly vast. The IBT is intended to serve as a reminder that all of the pieces of human capability must all fit together.

The notion of everything fitting together raises what I have come to call “the jigsaw puzzle problem.” The analogy here is that the whole cognitive science endeavor has as one of its goals the construction of a complete and detailed picture of human cognition and performance (which would nicely support the interactive behavior triad). There are multiple ways one might go about trying to generate such a picture. The divide-and-conquer approach characteristic of the field has subdivided the picture into some large number of little pieces. Most individual researchers in the cognitive sciences, or more specifically, most theorists of Type 2 (Gray, chapter 1, this volume), have taken responsibility for different pieces. That is, they have taken on the task of figuring out what their part of the overall picture, and only that part of the picture, looks like.

Of course, if we want to understand the complex behaviors we see in the real world (e.g., pilots), what we need is a comprehensive theory that encompasses, subsumes, or includes many of these Type 2 theories, or what has been termed a Type 1 theory. The job of the Type 1 theorist, then, is to take all the little jigsaw puzzle pieces and assemble them into a coherent picture. There are certainly multiple approaches one could take to this, and several have indeed been tried by different Type 1 theorists. One approach is to start with the idea that a central controller will be necessary and then gradually incorporate more and more of the Type 2 pieces into the monolithic controller. Type 2 theories will thus be subsumed into a larger system. This appears to be the Soar approach (Newell, 1990). A vaguely similar approach was taken with ACT* (Anderson, 1983) and early versions of ACT-R (Anderson, 1993), which treated declarative memory as a separate component but put all other functionality into the central controller. This produced a fairly monolithic architecture looking something like the one depicted in Figure 30.2.

This figure already looks dated by its simplicity. Indeed, the early 1980s architectures were almost entirely concerned with modeling central cognition and not perceptual–motor activities and thus such capabilities were not included in these architectures (though they perhaps could have been). I suspect that in the long run this monolithic kind of approach will not scale up to cover the entire picture very well and has the added drawback of not mapping easily on the more modular conceptualization of the brain offered by modern cognitive neuroscience.

An alternative approach, one represented by EPIC (Kieras & Meyer, 1997) and more recent versions of ACT-R (Anderson et al., 2004) is to embrace the notion of modularity and simply coordinate multiple modules with a central controller, rather than trying to do everything within the context of a central controller. This yields an architecture organized more like what is depicted in Figure 30.3. Given my research history, I unsurprisingly see this as a framework with a good chance of ultimately supporting the full picture. Not every architecture will include the same number of subsystems (neither ACT-R or EPIC currently include a module for emotion; different architectures are likely to disagree about the number of visual subsystems to represent, etc.), but this general organizational strategy still applies. The resemblance of this style of organization to the model human processor of Card, Moran, and Newell (1983) is not accidental, and this bears a meaningful similarity to Baddeley’s (1986) well-known conceptualization of working memory, with the controller in the “central executive” role. Thus, despite the relatively recent emergence of EPIC and the current organization of ACT-R, this is neither a new or especially radical conceptualization.

Another salient point is that several modules (e.g., vision) in this scheme have a more or less direct...
connection to the external environment, though that is not depicted in Figure 30.3. I chose to omit it from this diagram simply to note that some architectures may give much higher priority to understanding the internal components than to the role of interaction with the external world. Of course, the interactive behavior triad suggests that this would not be an optimal approach, but reasonable theorists could disagree on the centrality of such a connection.

Alternately, there are camps within cognitive science that eschew the notion of a central controller and see global control as an emergent property of a modular organization (e.g., the ICS system of Barnard, 1999). Figure 30.4 is a general representation of how such an architecture might be organized. Such a decentralized architecture is probably more representative of connectionist approaches such as Atallah, Frank, and O’Reilly (2004), though there are some interesting symbolic/connectionist fusion approaches which have something of this feel (e.g., Just, Carpenter, & Varma, 1999). Approaches of this kind are interesting and definitely have merit, but I have yet to see any such system scaled up to a task as intricate and complex as piloting. Whether it is ultimately possible is still an open question.

Again, the inclusion of connections to the external world is not intrinsically a property of this kind of organization; I am simply illustrating how one might construct such a system. I do generally agree with the notion that the environment plays a key role in assuring good overall control flow in the human cognitive system, but I do not intend to claim that systems without central controllers are better suited to addressing such issues.

I am not suggesting that other organizations are impossible or unlikely. My goal here is not to produce a catalog of all styles of cognitive architecture, but to point out that there are multiple approaches for integrating various Type 2 theories into a larger Type 1 architecture. Unfortunately, this is happening on only a very small scale; cognitive scientists who are engaged in building Type 1 theories are few and far between. On the basis of this analogy, the task here should be somewhat like solving a jigsaw puzzle; simply fit together all the relevant pieces, and then the picture should be complete. If it were that easy, I suspect more researchers would be engaged in this task.

Of course, it is in fact much more complex than this. Why? Why are the theorists and developers behind Soar or ACT-R or EPIC unable to simply assemble a collection of pieces that have been generated by Type 2 theorists? As someone who has tried to do this, I think there are multiple reasons. The first problem is that Type 1 theories are generally designed to support modeling and are thus executable simulation systems. Why is it that Type 1 theories tend to be computational
is itself an interesting issue, but it is beyond the scope of the current discussion; take it as a given that most Type 1 theories are computational or at least computationally oriented. To produce such a system, all of the pieces of the system must be specified in enough detail that an implementation is possible.

Lack of complete specification is, in fact, not unusual in cognitive science. Of course, there have always been exceptions and some things cleanly quantified; in fact, some quantitative and implementable formalisms describing certain aspects of behavior have even been elevated to “law” status, such as Fitts’ law for aimed motor movements (Fitts, 1954) or the Hick-Hyman law for choice reaction (Hick, 1952; Hyman, 1953). Unfortunately, these truly have been the exception rather than the rule. Returning to the jigsaw puzzle analogy, assembling a collection of informal and underspecified theories into a coherent whole is essentially impossible; it would be like trying to piece together a collection of amorphous, shifting amoebae. The good news is that this is changing, as several of the chapters in this volume clearly demonstrate. But it is not clear that there are yet enough well-specified Type 2 theories to assemble a collection of them into a meaningful Type 1 theory. I suspect—or at least hope—that this situation will change substantially in the next few decades.

Unfortunately, even if the cognitive science landscape were populated with well-specified and empirically convincing Type 2 theories, those of us in the business of constructing Type 1 systems would still face an enormous uphill battle in trying to assemble them. Consider the perspective of a Type 1 theory builder. A general organization is chosen, and there is some idea about what the overall picture might look like. The landscape of Type 2 theories is surveyed and the builder then goes out in search of puzzle pieces that can be fit together to form an overall picture. The unfortunate reality of the situation is that right now, the chance that the right pieces will all be available and will actually fit together is quite small.

Why? Because unlike real jigsaw puzzles where a picture is cut up one time in such a way that all pieces fit with something else, in the cognitive sciences there has been no central oversight of how the pieces are cut. It is as if everyone is given access to an extremely

![Fig. 30.4 Modular architecture with no central controller (distributed control).](Gray_CH30:Gray 1/12/2007 1:04 PM Page 436)
low-resolution version of the total picture, allowed to cut out one’s own piece, and then asked to work on improving the resolution of just that piece.

This leads to three closely related and challenging problems. The first is what I call the piece identification problem, the second the piece fit problem, and finally the piece distribution problem. With the piece identification problem, researchers not only get to pick out the part of the picture to call their “piece,” but they also get to choose the label for that piece. In building a Type 1 theory, in an attempt to have broad coverage, one would ideally want one piece from each area of the picture. But it is hard to know what part of the picture a piece represents because there is such little uniformity in terminology or input/output relations. Terminology is probably the easiest to illustrate. Considering all the literature that contains the term, it seems apparent that any Type 1 theory is going to have to address phenomena associated with “attention.” But what is attention? According to various Type 2 researchers in vision, we have selective attention (Broadbent, 1954), divided attention (Triesman & Gelade, 1980), object-based versus location-based attention (Duncan, 1984), and attentional sets (Folk, Remington, & Johnston, 1992). In central or cognitive work on attention, there is the limited attention in dual tasking (Pashler, 1994), attention given to various features in categorization (Kruschke, 1992), the attentional blink (Broadbent & Broadbent, 1987), and the attentional capacity that is freed when tasks are automatized (Shiffrin & Schneider, 1977). Not only are there many different senses of attention, but there is little contact and even less unification between various senses. So even identifying which pieces to try to fit together is a challenge for the Type 1 theorist. This is the most obvious of the three problems, so I will not elaborate further.

The piece fit problem has to do with input/output relations. Because Type 2 work has become so compartmentalized, there is often little or no discussion or specification in Type 2 theories of where the inputs to a particular mechanism come from and what form they take, nor where the outputs go and what form they take. Take, for example, models of visual search, of which there are many (see chapters by Wolfe [chapter 8] and Pomplun [chapter 9] in this volume for examples of solid Type 2 visual search theories). Most (but not all) models of visual search stipulate that somewhere in the connection to the visual search subsystem (or visual selective attention) there exists some information about the target being searched for, for example, that it is red. But what these models leave unsaid is what can be specified or how it is specified. In the typical visual search experiment, objects in the visual field appear entirely at random (or random from among a fixed set of locations). Experimental participants typically know either nothing in advance about the location of the target or they have specific information about the target location (e.g., from a cue), which may or may not be accurate. Assuming the visual search system is at least sometimes driven in a top-down fashion, what can the “top” module specify? Can approximate location be specified? People often have vague but meaningful expectations about where things will be located. For example, when searching a Web page for a link, people can reasonably expect that the target should be “somewhere on the left.” Can that kind of expectation be passed to the visual selective attention system, and if so, how?

The situation is not much better on the output side. Some more coarse models predict only overall response times for present/absent judgments. Other models generate simulated movements of visual attention and/or point-of-gaze unfolding over time. The piece fit problem, then, is that there is no simple way to gather Type 2 theories or models and match them up with other system components.

I suspect that the extent of this problem is grossly underestimated by most cognitive scientists, so I want to provide a somewhat detailed illustration of the problem. To do so, I want to draw on my experience in constructing the perceptual–motor subsystems for ACT-R. At the time I began working with ACT-R, it was a monolithic single-channel architecture focused primarily on central cognition (see Figure 30.2). The good news was that some useful work on giving ACT-R elementary visual and motor capabilities had already been done, under the ACT-R visual interface described in Anderson, Matessa, and Lebiere (1997). However, that system was entirely single channel; that is, if the motor system was busy moving the hands, then everything else in the system stopped. This may be a reasonable approximation in some contexts, particularly tasks where the bulk of the task is pure cognition, but that approach is inadequate for multiple-task situations or highly interactive domains where cognition, perception, and action are all interleaved. The power of this kind of interleaving has been apparent for some time, as laid out in the model human processor and as clearly demonstrated by the model human processor–based models of telephone operators presented in Gray, John, and Atwood (1993).
At the time I began my work on ACT-R, the EPIC architecture, which has a modular organization and thus allowed for parallel cognitive, perceptual, and motor activity, was just beginning to attract attention. Our first thought on how to modularize ACT-R was to simply replace the “cognitive processor” in EPIC, also a production system, with ACT-R’s production system. We knew that a similar effort was under way using the Soar architecture as a cognition alternative (e.g., Chong & Laird, 1997). After considerable analysis of EPIC, I decided that this strategy would work or less work for some modules (particularly motor and vocal modules) but that it could not work for others, particularly vision. What I determined was that the input/output properties of the cognitive systems were different enough that the interface between vision and cognition had to be fundamentally different for ACT-R and for EPIC’s cognitive processor. Why? EPIC’s philosophy of minimal commitment to assumptions, which is a powerful research strategy, allowed a great deal of flexibility in how vision and cognition could interact in EPIC. ACT-R, however, had already made a great many long-standing commitments on numerous fronts, especially with respect to memory. This meant that the communication between vision and cognition had to be much more constrained.

A more concrete description of one of the issues may help clarify. In EPIC, the visual system deposits representations of the visual scene (e.g., working memory elements representing visible objects) directly into the production system’s working memory. When anything changes (for example, a new object appears, or something as mundane as an eye movement rendering new objects visible and old ones not), new representations are added and, if necessary, old ones simply deleted. This seems straightforward and obvious. However, it also happens not to work for ACT-R. Working memory elements in ACT-R have more regulated structure than the arbitrary lists used by EPIC, but more importantly, they have interactions with other parts of the system. They have activations that decay, strengths of association with other items, and counts of times they have been accessed to compute base-level activation. Simply deleting such elements is not straightforward in ACT-R.

In fact, what we ended up with in ACT-R is a system where the vision module does have an EPIC-like representation where things are deleted and added as necessary; this store is visual iconic memory. But unlike in EPIC, contents of this store cannot simply be copied into declarative memory and deleted when no longer current; in ACT-R, gatekeeping between the iconic store and declarative memory is provided by an attentional mechanism. No such mechanism is present in EPIC, which has no need for a “covert attention” construct. The EPIC design is somewhat cleaner and definitely more parsimonious, but it does not provide for things like a decaying memory for what was seen or associative learning of object locations over time.

The real point here is not to argue about which architecture has the better visual system; the real point here is that your theory of vision is constrained by the properties of your theory of cognition, and vice versa. But since most Type 2 vision theorists do not worry much about cognition, and often use tasks in their research that are carefully designed (either implicitly or explicitly) to minimize the role of cognition, their theories will not reflect such constraints. The same applies to many Type 2 cognition theorists as well, but of course visual constraints are being neglected. Thus, even when it is possible to find two puzzle pieces that cover adjoining sections of the picture, they are unlikely to smoothly fit together.

In contrast to the piece identification and piece fit problems, which occur because of how Type 2 research is done, the piece distribution problem occurs because of where Type 2 research is done. In trying to collect as many pieces to assemble as possible, a Type 1 theory builder would find that certain parts of the picture have many pieces, many of which overlap, thus creating a small pile of pieces. However, there would also be portions of the picture with no pieces whatsoever. There are large areas that are simply not covered. Some of these are understandable, while some of them are quite surprising. To take an example from my conversations with Dave Kieras, there is the issue of retinal availability functions. We know, for instance, that color perception is “best” in the fovea and “worst” at higher eccentricities. What we do not know is the shape of the function mapping eccentricity to availability for numerous visual features such as size and orientation. It is striking that this kind of basic information about the perceptual system is not readily available in the research literature.

To again take an example from my own research, consider the two displays depicted in Figure 30.5. In my laboratory participants are trained to execute relatively simple routine procedures on each of these simulated devices. Critically, each of these procedures is isomorphic in terms of the goal–subgoal–action
hierarchy. That is, both tasks contain the same number of subgoals and steps. For example, in the “phaser” task, the first subgoal is to charge the phaser. To do this, subjects click the power connected button, then the charge button, then they wait for the charge indicator (the bar on the right) to fill, then click stop charging, and finally click power connected. The “transporter” task starts with a subgoal with the same structure but with a different name and using different buttons. Both tasks have the same number of subgoals, and each subgoal has the same number of steps. Subjects are required to practice the tasks to an accuracy criterion so these tasks are essentially routine procedures.

An approach for analyzing and predicting behavior in such tasks is GOMS analysis (goals, operators, methods, and selection rules; Card, Moran, & Newell, 1983). GOMS models have an extensive history and strong empirical track record of successfully capturing human behavior for routine procedural tasks (see John & Kieras, 1996, for a review). Thus, these tasks should already be covered by an extant account, because according to a GOMS analysis, these tasks should be equivalent (other than some minor differences in mouse

FIGURE 30.5 (a) Display for the “phaser” task. (b) Display for the “transporter” task.
pointing time based on Fitts’s law). Because of this, in my lab we refer to these tasks as “GOMS-isomorphic.”

In fact, human performance on these tasks differs substantially in terms of time to execute individual steps of the procedures and error rates at each step. For example, the error rate on the first step differs by more than a factor of three between the two isomorphs. (For more details on this, see Byrne, Maurier, Fick, & Chung, 2004). What is obviously different between the two displays is the layout of the controls, clearly something in the visual domain but which may affect how the controls are functionally characterized by the cognitive system. However, there is essentially nothing in either the visual attention literature or the literature on control of sequential behavior that predicts how these two tasks will be different or explains why. This is the kind of gap alluded to earlier. While this particular gap is on the boundary of vision and cognition, I strongly suspect that there are a preponderance of similar gaps along other boundaries between other subareas.

This problem is not simply getting the input/output relations between two puzzle pieces to match up; there are important parts of the overall picture for which no pieces exist. Additionally, there are other parts of the picture, such as how people perform analogical mapping or supervised category learning, for which there are multiple available pieces. Obviously, this makes assembling the pieces constructed by Type 2 theorists a difficult task, as anyone who has ever tried to solve a difficult jigsaw puzzle with a substantial number of missing pieces and many nearly identical ones can attest.

Between these three problems, one might be inclined to conclude that building comprehensive Type 1 theories is impossible and that realizing an even more complete view of human performance like the interactive behavior triad is simply a pipe dream. While I think the challenges are substantial, I do believe the last decade has seen enormous progress on this front. Theories such as ACT-R and EPIC continue to push boundaries, and this volume positively indicates the existence of such a thing might encourage Type 2 researchers and might lend credence to the idea that the Type 1 theorists are actually interested in covering areas outside of their traditional specialty.

So, in the spirit of further progress, what can be done to help address this problem? Many things could be done by both Type 1 and Type 2 theorists to facilitate progress on assembling the big picture. I do not expect a sudden burst of productivity as a result of these suggestions, but rather I hope both groups will occasionally take these considerations to heart. The first issue is accessibility of the Type 1 architectures. Accurate or not, these architectures are widely perceived as monolithic, complex, and unapproachable. It is unlikely that this notion can be easily dismissed in the context of journal articles. To address this, Type 1 theorists need to do a better job of outreach. Material describing the architecture and how to think about it needs to be clear and approachable. Having worked with several Type 1 architectures, it is fair to claim that none have scored very well on usability for novices, which potentially could be helped by more GUI-based software tools to not only develop models but also to test and understand a model’s behavior as it runs. Good documentation is also crucial but often lacking.

Another recommendation would be if Type 1 theorists were explicit about what pieces their systems lack and why extant Type 2 models in those areas have not been incorporated. Admitting that one’s architecture is incomplete can be difficult; however, it may prompt Type 2 researchers to realize that their piece is needed to fill an existing gap or to consider how their model might fit in with the larger picture.

Finally, the software architecture of most Type 1 systems do not lend themselves to inclusiveness. Most of these systems are relatively closed, and the code base is indeed monolithic. Obviously, modular architectures are likely to have an advantage here, since Type 2 theories can more easily be incorporated into a modular system. However, this is not enough. There should be well-documented rules for intermodule interaction—even an application program interface (API)—so that it is clear how a new module could be incorporated. The existence of such a thing might encourage Type 2 researchers and might lend credence to the idea that the Type 1 theorists are actually interested in covering areas outside of their traditional specialty.

Alternatively, what could Type 2 theorists do? This is a more difficult question because it raises another problematic issue. While it is relatively obvious why Type 1 system builders (especially those interested in comprehensive frameworks like the interactive behavior triad) would want assistance from Type 2 theorists, what is the incentive for Type 2 researchers? Unless journal editors start mandating that experimental tasks require more integrated performance or that theories and models have clear connections to a larger cognitive
system, change here is likely to be slow. The psychology laboratory, by its nature, tends not to be a place that engenders complex tasks that require diverse capabilities because getting experimental participants up to speed on tasks of high complexity is expensive.

However, most (or at least many) Type 2 researchers are honest with themselves; they realize that their domain is but a part of a larger picture and are at least open to trying to understand the whole of cognition and performance. While most Type 2 theorists might not be interested in pursuing such work themselves, many can at least see that such endeavors have scientific value. So, for those Type 2 researchers who are interested in supporting such activities, several things could be done to help Type 1 theorists synthesize Type 2 models into a more comprehensive architecture. (Note that these are intended as suggestions, no matter how much this may sound like a list of demands.)

Type 1 researchers would be helped if Type 2 theorists elucidated exactly what parts of the overall picture are and are not covered by each Type 2 model. The idea is to try to locate the puzzle piece within the larger picture and, just as important, to try to provide information about its shape so it is clear how it might be fit together with other pieces. Saying that “this is a signal detection model of selective visual attention” or “this is a diffusion model of decision-making” is not enough. Explicit context, delineation, and specification are needed. For example:

- What is and what is not included in this account? For those things that might have been included but were not, why were they not included? That is, was it theoretical, technical, pragmatic? Type 2 theorists often know a lot about where not to go in a domain, and Type 1 researchers would benefit from such wisdom in not repeating mistakes.

- What cognitive–perceptual–motor function does the model address? How would that be used by a larger system? What kinds of larger tasks would contain that function as a subtask?

- What does the model take as input and in what form does it take it? Does the model take inputs strictly from perception or from other cognitive subsystems? If the model takes perceptual input, what aspects of that input have to be preprocessed and how? What is the representation over which the model works?

- What does the model output? Products of the process (e.g., decisions), or timing, or both?

Does the model produce intermediate products (e.g., saccades) or a final time for the trial (or problem or whatever is being modeled)? Are there other subsystems (particularly motor) which are supposed to receive output form the model? If so, what information is passed on and in what form?

Unfortunately, many of these things obviously do not easily fit into journal articles. However, provision of such information (perhaps on a Web site) would make it substantially easier for Type 1 system architects to incorporate the relevant ideas, which has the potential of providing wide distribution of the ideas and thus increasing research impact. And who doesn’t want to increase their research impact?

Increasing impact is a goal that serves all members of the community. When Type 2 theories can be integrated into Type 1 systems and those systems improved by the result, everybody wins. The Type 2 theorist wins because their work will be seen by an audience beyond their particular specialization. Furthermore, the Type 2 theory will be not merely disseminated within the Type 1 community, but also it will have a chance to affect new tasks and new environments. The Type 1 theorist wins because the theory becomes more complete and, it is hoped, more accurate with a wider range of applicability. Broader and higher fidelity Type 1 human performance theories support stronger IBT-based analysis and modeling and thus should have significant potential for application to highly interactive real-world domains like aviation and medicine.

Thus, the effort involved in uniting Type 1 and Type 2 research is worth making, as the long-run payoffs should be substantial. The work in this volume represents an important step in the direction of this payoff, so I hope further workshops and volumes that reflect these concerns will become increasingly common. While cognitive science has not entirely avoided falling into the trap forecast by Newell some 30 years ago, if enough researchers are willing to commit to the endeavor, we have the potential to successfully build integrated theories.

Acknowledgments

I would like to thank the Office of Naval Research for its support under Grant N00014-03-1-0094 and the National Aeronautics and Space Administration for its support under Grant NDD2-1321. The views and
conclusions contained herein are those of the author and should not be interpreted as representing the official policies or endorsements, either expressed or implied, of NASA, ONR, the U.S. government, or any other organization.

References


formance with application to human-computer interaction. Human-Computer Interaction, 12, 391–438.