Mental Action and the Threat of Automaticity

Wayne Wu

> [I]nput mechanisms approximate the condition often ascribed to reflexes: they are automatically triggered by the stimuli that they apply to…It is perhaps unnecessary to remark that it does *not* seem to be true for nonperceptual cognitive processes…we have all the leeway in the world as to how we shall represent the objects of thought (Jerry Fodor (1983), p. 54-5).

> The movement of the natural causality of reason (practical reason in this case) to its conclusion in choice or decision is lived (by some) as action when it is really just reflex; distinctively rational reflex, to be sure, but not in any case a matter of action (Galen Strawson (2003), p. 244).

1. Introduction

The starkly opposed assertions of Fodor and Strawson highlight one controversy addressed in this essay: when does something count as a mental action? This disagreement, however, points to a deeper controversy, one intimated by the appeal to reflex as a contrast to genuine action. Reflex, such as blinking to looming visual stimuli or withdrawing one's hand from a burning surface, is a paradigm form of *automatic* behavior. As we shall see, automaticity is what makes decisions about mental agency controversial, but it has also in recent years led to unsettling conclusions regarding our conceptions of agency and agency itself.

Psychological research on automaticity reveals it to be a pervasive feature of human behavior. John Bargh and Tanya Chartrand (1999) speak of the "unbearable automaticity of being", arguing that "most of a person's every day life is determined not by their conscious intentions and deliberate choices but by mental processes that are put into motion by features of the environment and that operate outside of conscious awareness and guidance" (462). Relatedly, what is unbearable is that we are also zombies (in a non-flesh eating sense, thankfully). Francis Crick and Christof Koch (2001) have written about

1

"zombie agents" understood as "systems [that] can deal with certain commonly encountered situations automatically" where automaticity implies the absence of conscious control. But it is not just systems that are zombie agents, but subjects as well. They ask: "Could mutation of a single gene turn a conscious animal into a zombie?" Yet if our behavior is permeated with automaticity, aren't we all zombie agents?

The starkest consequence, however, is that automaticity appears to eliminate agency. Even if agency were unbearably automatic or zombie-like, it would still be agency. Yet I shall show that a common assumption regarding automaticity suggests that it is incompatible with agency. If so, it is not that agency is unbearably automatic or zombie-like. It isn't agency at all. And if there is no agency, there is *a fortiori* no free, rational, moral or conscious agency. To illuminate these issues, automaticity and its correlate control must be incorporated in a theory of agency. This essay shows how. The result is that we can acknowledge two seemingly inconsistent claims: that automaticity implies the absence of control and that agency, as an agent's exemplification of control, involves, and often requires, much automaticity.

What then is automaticity? In section 2, I review empirical conceptions of automaticity where psychologists came to reject a *simple connection,* namely that automaticity implies the absence of control. Yet philosophical reflections on mental agency also suggest that we should reject the simple connection, and in section 3, I develop an argument that suggests that adhering to it eliminates not only mental agency but agency *tout court.* This is the threat of automaticity. In response, I explicate the causal structure of mental agency in section 4, and then defend the simple connection in section 5 by embedding it in that structure. The final sections put these results to work: section 6 defuses the threat from automaticity, section 7 responds to the striking philosophical

disagreements about basic cases of mental action, and section 8 reflects on inferences from automaticity to claims about agency in cognitive science.

2. Automaticity and Control in Cognitive Science: A Very Brief and Selective Overview

The psychological literature on automaticity and control is voluminous, and as things have become complicated, I make no claims to completeness in what follows. Rather, I shall highlight specific features of the discussion of Schneider and Shiffrin (1977), work that in many respects set the terms of the empirical debate. My goal is to highlight how psychologists moved away from Schneider and Shiffrin's *simple connection* between automaticity and control.

Schneider and Shiffrin focused specifically on the notion of an automatic *process* within a familiar picture of the cognitive system as a set of connected nodes (a neural network) where processing involves sequential activation of a subset of nodes. Consequently:

> An automatic process can be defined…as the activation of a sequence of nodes with the following properties: (a) The sequence of nodes (nearly) always becomes active in response to a particular input configuration, where the inputs may be externally or internally generated and include the general situational context. (b) The sequence is activated automatically without the necessity of active control or attention by the subject (2).

The simple connection is that automaticity implies *the absence of control (or attention) by the subject.* Automaticity is then clarified by explaining control, and on this, Schneider and Shiffrin note that "[a] *controlled process* is a temporary sequence of nodes activated under control of, and through attention by, the subject" (ibid.). Control, as they conceive it, involves the deployment of attention. Thus, the initial foray into defining automaticity and

control relies on two links: the simple connection relating automaticity to the absence of control and the conception of control in terms of the direction of attention.

Things rapidly became more complicated. John Bargh (1994) notes in a review of the social psychology literature on automaticity that theories after Schneider and Shiffrin gravitated to four factors as defining automaticity: automatic processes are "unintentional, occur outside of awareness, are uncontrollable, and are efficient in their use of attentional resources" (2). The problem, as Bargh emphasizes, is that many paradigm cases of automaticity fail to exemplify all four properties. Shiffrin later observed that "there do not seem to be any simple defining features of automatic and attentive processes that can be applied in complete generality" (1988, p. 765) and he identified *ten* typical contrastive features of automatic and controlled processes (in a communication, he has indicated giving up on providing necessary and sufficient conditions for automaticity). Similarly, Thomas Palmeri (2002), in an encyclopedia entry on automaticity echoes Bargh's characterization, writing that: "Automaticity refers to the way we perform some mental tasks quickly and effortlessly, with little thought or conscious intention. Automatic processes are contrasted with deliberate, attention-demanding, conscious, controlled aspects of cognition" (290). He then goes on to list *13* contrasting features between automatic and controlled processes (table 1, 291). Schneider (2001), in his entry on "Automaticity" for the *MIT Encyclopedia of Cognitive Science* notes that "[a]utomatic processing shows seven qualitatively and quantitatively different processing characteristics relative to controlled process" (63). Current psychological conceptions of automaticity have clearly gone far from the simple connection.

Why have psychologists given up the simple connection? Gordon Logan has observed (1988), given the characterization of control as the (generally conscious)

deployment of attention, the simple connection thereby predicted that automatic phenomena would be freed of the constraints imposed by attentional processing such as capacity or load limitations.[1] The problem, Gordon emphasizes, is that empirical work has shown that many putatively automatic phenomena often are subject to attentional limitations or are otherwise influenced by how attention is directed. Psychologists have continued to construe control to be or at least to involve deployment of attention, so the way to accommodate evidence of capacity limitations in automatic processes is to sever the simple connection (Gordon (1988) analyzes automaticity in terms of memory via his *instance theory*).[2]

In what follows, I do the reverse: I retain the simple connection and give up equating control with attention although on my view, control *in action* implies the deployment of attention. This will allow us to address the consequences of automaticity on agency noted above. The most serious of these is the threatened loss of agency in the face of automaticity. I turn to this threat now which is most apparent in mental action.

3. Mental Ballistics

Given the armchair nature of their work, it is striking that philosophers have largely neglected mental actions.[3] Moreover, when philosophers speak of mental actions, there is striking discord as to what counts as an instance. For example, you ask me, "Who was the prime minister of Czechoslovakia when the Soviet Union invaded?" and I try to recall.

---

[1] It is worth pointing out that a canonical type of automatic process is attentional, namely attentional capture, and that a function of attentional capture is to disrupt other activities.

[2] In the case of mental action, I will emphasize the central role of memory in many forms of cognitive attention and this resonates with certain key aspects of Logan's theory. In contrast to his account, however, I emphasize the simple connection.

[3] The essays in Soterious and O'Brien (2009) are a corrective to this neglect.

When I recall the answer, I then judge, "It was Dubcek."[4] Here we have a remembering and a judging, two mundane mental events. Are these actions? Christopher Peacocke claims that they are. On the other hand, Alfred Mele (2009, 19) asserts that remembering is never an action and Strawson denies action status to both. Who is right?

In contrast, similar questions regarding mundane bodily events, say the grabbing of an object when one is asked for it, elicit broad agreement as to their status as actions. I conjecture that the disagreement stems in part from the fact that mental actions are typically automatic and that there are differing intuitions about the import of automaticity with respect to agency. Work in social psychology has cataloged the pervasiveness of automaticity even in goal directed behavior where control is conceived as delegated to the environment, obviating any need for conscious control on the part of the subject (Bargh and Ferguson (2000)). Bernhard Hommel (2000) has spoken of intention as a "prepared reflex", the upshot being that intentions allow the environment to take control of behavior without the need of further intervention (control) by the subject.

I interpret Strawson's arguments against mental action as exploiting the ubiquity of automaticity. Strawson argues that most of what we take to be mental actions involve mental ballistics, namely things that happen automatically. Strikingly, he includes deliberation and imagination. A natural reaction to this constriction on the sphere of mental action is to take it as a *reductio* of Strawson's assumptions, but his arguments provide an opportunity to focus on the significance of automaticity in action.

To bring out the threat, let us push Strawson's argument to its logical conclusion. Speaking of imagination, Strawson observes:

> When one has set oneself to imagine something one must obviously start from some conceptual or linguistic specification of the content (spangled pink elephant), and

---

[4] The example is from (Peacocke 1998).

given that one's imagining duly fits the specification one may say that it is intentionally produced. But there isn't intentional control in any further sense: the rest is a matter of ballistics, mental ballistics. One entertains the verbal specification and waits for the mechanism of imagination—the (involuntary) spontaneity of imagination—to deliver the image (241; my emphasis).

Why should there be control in any further sense than the one Strawson specifies, namely that the process fits what one intended? I shall return to this, but whatever this further element is supposed to be, Strawson thinks it is incompatible with the automaticity of imagination. In imagination, we have only mental ballistics, some automatically generated process for which we can only wait once we have specified the relevant parameters, say the image of one's mother's face or of a pink elephant. Thus,

When one sets oneself to imagine anything there comes a moment when what one does is precisely to *relinquish control*. To think that the actual content-issuing and content-entertaining that are the heart of imagining are themselves a matter of action seems like thinking, when one has thrown a dart, that the dart's entering the dartboard is itself an action (242; emphasis added).

Similarly for practical deliberation:

Very often [in the case of practical deliberation] there is no action at all: none of the activation of relevant considerations is something one does intentionally. *It simply happens*, driven by the practical need to make a decision. The play of pros and cons is *automatic* —and sometimes unstoppable (243; emphasis added).

So, deliberation and judgment are not actions and other cases like memory recall will be more grist for Strawson's mill. It would seem that there is no space for mental agency.

Still, Strawson notes that "[t]here is of course such a thing as mental action" (231) and finds space for it in *stage setting*:

the role of genuine action in thought is at best indirect. It is entirely prefatory, it is essentially—merely— catalytic. For what actually happens, when one wants to think about some issue or work something out? If the issue is a difficult one, then there may well be a distinct, and distinctive, phenomenon of setting one's mind at the problem, and this phenomenon, I think, may well be a matter of action. It may involve rapidly and silently imaging key words or sentences to oneself, rehearsing

inferential transitions, refreshing images of a scene, and these acts of priming, which may be regularly repeated once things are under way, are likely to be fully fledged actions (231).

The problem is that he cannot avoid the threat posed by automaticity. What is it that goes beyond mental ballistics in stage setting? Strawson speaks of "setting one's mind at the problem", but what is this setting? "[I]t may involve rapidly and silently imaging key words or sentences to oneself, rehearsing inferential transitions, refreshing images of a scene". Yet surely imaging, rehearsing inferences and refreshing images are just instances of the kinds of cases his previous arguments ruled out as actions, namely imagination, deliberation and recall. Thus, he cannot avoid the threat even in the remaining space he allows for mental agency.

To be fair, Strawson points out other cases of mental action: shepherding or dragooning back a wandering mind, 'concertion' in thought, stopping, piling and retaking up thoughts that come to us quickly, or an active receptive blanking of the mind (231-2). All of these are catalytic, and mental action goes no further than this. Yet he also says that these need not be actions, and then their occurrence would be automatic. But if automaticity is found in these cases too, then is there any missing agentive element to be found here that we failed to find in deliberation, imagination or recall? As we have seen, in each case Strawson considers, the uncovering of automaticity in mental ballistics leads to a negative answer. The threat of automaticity is ubiquitous, and accordingly, action seems to be eliminated from the mental sphere.

Strawson explicitly invokes the concept of automaticity at only one point, but the idea is captured in his talk of reflex, ballistics, and mere happenings. There is in all of these a contrast with the agent's being in control. This is just an instance of the simple connection that originally guided contemporary psychological conceptions of

automaticity: where x is automatic, then x is not a case of control. If we think of action just as an agent's exerting control, automaticity implies the loss of agency.

The threat extends to *bodily* agency. Consider moving one's arm, say to reach for a glass. The specific movement that the agent is said to make is individuated by certain parameters, say the specific trajectory, speed, acceleration, grip force, sequence of muscle contractions, angle of the relevant joints and so forth. Yet in a similar sense, the production of these features is itself merely automatic (and indeed, the initial movement of the arm in reaching is literally ballistic). Given that any token movement is individuated by automatically generated features, one might argue that the concrete movement itself is automatically generated. After all, what features are left for the agent to control?

Even in bodily action, it seems that we are pushed back to talk of stage setting, but there is no avoiding the threat there. After all, intentions set the stage for intentional bodily movement, but Strawson's argument that deliberation and its product, intention, are essentially mental ballistics shows us that agency evaporates here as well. So, the threat from automaticity is total: mental and bodily actions disappear. This conclusion is more extreme than Strawson allows, but it seems unavoidable given three initially plausible claims: (1) the pervasiveness of automaticity in human activity, (2) the conceptual connection between actions and an agent's exertion of control, and (3) the simple connection linking automaticity to the absence of control. I take it that the loss of all action is a *reductio* of these claims. Since (1) is empirically established and (2) a conceptual truth (on my view), the source of the problem is the simple connection.

It is *prima facie* plausible that the simple connection must go, for automaticity seems required in skilled action. In many cases where we practice an action, we aim to make that exercise automatic yet at the same time do not aim to abolish our agency. Thus,

when I practice a difficult arpeggio in the left hand on the piano, my aim is to master a particular passage so that I can play the entire piece correctly. Getting the left hand to perform automatically is precisely what is needed to play the piece as I intend. Indeed, with the left hand "automatized" I focus on the right hand's part, one that requires effort and attention. Playing with both hands, one automatically and the other with explicit effort, I play a passage in the piece intentionally. I think this is right, but in what follows, I show that we can also hold on to all three claims that generated our *reductio*. To see this, we must embed automaticity within a theory of agency.

4. Mental Action and Attention

To start, let us warm up with a few mental exercises (please go along): (1) Imagine your mother's face; (2) Recall the capital of Britain; (3) Add 2+2; (4) (In your head, answer): Is it wrong to torture a kitten?

I am going to conjecture that for most of these actions, the required answer simply "popped in your head". At first glance, what you did in each case was to undertake a mental action—you did something in your head—and in each, the desired product comes to mind seemingly automatically. This is the data of automaticity on which Strawson's argument rests. In this section, I outline a causal account of mental action, one in which intentions play a specific role in organizing and generating action. Central to this conception will be attention, and indeed, I argue, borrowing from William James, that conscious mental action is cognitive attention.

Our primary target will be those cases where there is something it is like to act in a certain way mentally, and thus, we are interested in conscious mental actions. I will speak of such cases as bringing thought contents to awareness, whether this involves recalling

some content, imagining some content, internal speech, reasoning via transitions in contents, and so on with the limiting case that of keeping hold of a specific thought in the fore in one's cognitive awareness. In many cases, prior to bringing the content to awareness, that content was recorded in the subject's memory, where memory is broadly construed so as to involve any content retained on the basis of alterations to the mind in light of past activity (including perception of the environment or synthesis of content already stored). The reason for holding conscious mental actions as so based in memory is that our thoughts are not created *ex nihilo,* but depend on prior thoughts or perceptions. Thought contents that we bring to awareness are either previously encoded (e.g. in episodic recall) or synthesized from previously encoded memories (e.g. as when we imagine fantastical creatures like unicorns). This appeal to a dependence on memory acknowledges that mental actions can also depend in analogous ways on the deliverances of perception (as when we choose some object perceived to think about). This does not affect the following points, so for simplicity, I shall largely not mention perceptual-based thoughts.

Once we recognize that the mental actions at issue, generically bringing thoughts to awareness, are based in memory, we immediately confront a selection problem. There is more than one item stored in memory, but conscious thought has to be specific. We do not activate all memories, there being limits to our capacity as to what we can entertain. This is an empirical point, but given these limitations, the mental action at issue requires selection of memory. This leads to a Problem of Selection: which memory should be brought to awareness?

For example, a mathematics graduate student intends to recall some mathematical proposition during an exam. She is motivated to recall a certain axiom of set theory, but of

course, she has in memory a huge number of mathematical propositions. Nevertheless, the relevant proposition occurs to her, and she can then use it in subsequent reasoning, which itself calls upon further dredging through memory. Action requires solving this Problem of Selection, specifically reducing a set of many possible inputs by selection of exactly what is relevant to the task at hand (similar points arise in respect of perceptually-based thoughts). For simplicity, we can speak of this as reducing a set of many possible inputs to a single relevant input.

A similar point arises regarding output or behavior. Returning to our mathematician, she has, in fact, many choices open to her even after selecting a specific memory. There are many responses that she can give: she can deploy the remembered axiom to construct her proof, she can think of how best to format a representation of that axiom for the final version of a journal article, or she can imagine writing a song where a statement of the axiom is used as part of a catchy jingle (and much more besides). There is a set of possible things that she can do with the thought in light of bringing it to awareness. In mental action, then, doing something is navigating through a *behavioral space* defined by many inputs and many outputs and their possible connections. Bringing a thought to awareness is selecting the relevant memory so as to inform a specific type of conscious mental activity. Mental action requires the selection of a path in the available behavioral space. In earlier work, I have called this Problem of Selection the *Many-Many Problem*. The Many-Many Problem is, I have argued, a metaphysically necessary feature of bodily action (2011), and it is exemplified in the mental case. Certainly, in actual mental actions, we have the Many-Many Problem, and that weaker claim will suffice for our purposes.

Solving the Many-Many Problem is a necessary condition on agency. But clearly not any "solution", namely a one-one mapping of input to output, will suffice. In the

context of an exam, our mathematician wants to find a solution to the mathematical problem she is tackling. She is not, at that moment, interested in writing songs about set theory or formatting a text for publication. Should such outputs—writing a song, attempts to format a text—be what results during the exam, she would see these behaviors as inadvertent and involuntary. Moreover, should the selections at issue routinely pop into her head in other contexts such as when she is discussing a poem (but not one about set theory!), then this would be something odd, inconsistent with her goals at that time.[5]

What this points to is that solving the Many-Many Problem cannot be *inconsistent* with one's current intentions. The content of one's current intentions sets the standard by which one's actions are successful or not, and the way to ensure consistency with intention is to require that solutions to the Many-Many Problem are not *independent* of intention. Dependence of selection on intention should then be understood as the causal influence of intention on selection, and this is intuitive: our mathematician recalls the relevant axiom in set-theory precisely because she intends to solve a problem in set theory of a certain sort; and she constructs a proof precisely because the problem at issue is to prove a certain theorem. More abstractly, the behavioral space that identifies the agent's action possibilities for a given time is constrained by intention such that a specific path, namely the intended one, is prioritized.

In mental action, solving the Many-Many Problem by making appropriate selection imputes a certain form of activity to the agent. This activity, however, is not an additional thing that the agent does so as to act in the intended way. Moreover, selection is not necessarily conscious. Rather, these are just aspects of or part of the mental action itself,

---

just as a specific movement of the fingers or perhaps the specific contraction of muscles in the hand is part of tying one's shoelaces. These are not additional things that one does so as to tie one's shoelaces nor does one need to be conscious of those features of the action. The same point applies to any selection relevant to solving the Many-Many Problem.

Earlier I mentioned the central role of attention in agentive control, and we can now see that the selection of a path in behavioral space that constitutes solving the Many-Many Problem in mental action yields a form of cognitive attention. Consider this oft-quoted passage from William James:

> Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others (James 1890, 403).[6]

I have found this passage often invoked when philosophers discuss attention although they typically go on to discuss *perceptual* attention (I am guilty of this as well). In this passage, however, James speaks not only of perceptual attention but attention in *thought*, what we can call cognitive attention. Accordingly, it is important to bear in mind a critical difference between attention in perception and attention in thought: generally, the inputs that must be selected in perception are typically simultaneous in perceptual awareness, but the inputs, namely thought contents, are not simultaneously given to awareness in thought.[7] Thus, when you perceive a cluttered scene, looking for your lost keys, vision simultaneously gives you multiple objects and visual attention is part of an active searching for that object among many actually perceived objects. In contrast, when one is trying to find the right thought (say the right axiom to construct a proof), one is not in fact cognitively aware of multiple simultaneous thoughts (i.e. encoded mnemonic content). The

---

[6] See also Peacocke op. cit., p. 70.
[7] Of course, for perceptually-based thoughts, the putative objects of thought can be given simultaneously.

14

thoughts from which we must select the relevant content are not actual objects of awareness in the way that perceived objects are, but (to borrow James' phrasing) only possible objects of awareness. They are the items that we have in memory, and thus, the Many-Many Problem in thought does not involve a conscious act of selecting the appropriate mnemonic items. James does in that passage speak of how it seems to one that one is confronted with multiple possible thoughts, but I read this as pointing to one's sense of multiple behavioral possibilities (recall the behavioral space).

The way I have put the point is that in bringing a thought to awareness, or as James says, taking possession of it by the mind, we have to select the relevant memory. This is just to solve the input side of the Many-Many Problem in bringing the thought to awareness. The solution to the Many-Many Problem, understood as path selection leading to awareness of a thought content fits with James' description of cognitive attention as the selection of a possible train of thought where focalization and concentration of consciousness, the awareness of a specific thought, are of its essence. Given James' characterization of attention, such conscious mental actions of bringing thoughts to cognitive awareness are Jamesian forms of attention. Plausibly, this category of conscious cognitive awareness (broadened to incude perceptually-dependent thoughts) exhausts the category of conscious mental action. So, all conscious mental actions are instances of Jamesian cognitive attention.[8] With this concrete proposal for mental actions, I return to the simple connection.

5. *The Simple Connection*

---

[8] There are, of course, different things we can mean by "attention". I am here emphasizing the insight in James' description, what he takes to be part of what we all know about attention. The general point is that action requires attentional selection given the Many-Many Problem.

I shall define automaticity in action in terms of the absence of agentive control and I show that a single process can be both automatic and controlled. How is this possible? The answer is a relativization that seems not to have been made in the empirical literature though it echoes GEM Ancombe's (1957) notion of action as *intentional under a description*. The account I shall give, however, is not concerned with descriptions of actions but rather properties of actions. Nevertheless, it links a traditional concern in philosophy, namely intentional action, with a traditional concern in psychology, namely automaticity. It is in this way that I propose to incorporate the latter in the former and respond to the threats and difficulties that automaticity brings.

We relativize automaticity and control in respect of what the subject intends to do, namely the types of actions, Φ, that she aims to bring about. We can then speak of control and automaticity in respect of Φ, for any action type Φ that the subject can be said to do.[9] As a consequence, we can affirm that automaticity is the absence of control *and* that an action event can exhibit both automaticity and control. The crucial element that bridges philosophy and psychology is that of solving the Many-Many Problem. In the definitions, I will speak of a subject's *executing* a solution to the Many-Many Problem when the subject acts in a certain way though typically in discussion, I will speak simply of solving the Problem, with the executing implicit.

My account assumes a notion of top-down causal influence where this depends on imposing a hierarchy in processing between cognitive and perceptual systems.[10] There are

---

[9] These Φ's are also the basis of descriptions under which the action can be said to be intentional. It is not clear that the definitions of automaticity or control of an event in respect of Φ to be given below are equivalent to claims about the intentionality of the event under the description "the Φ". Whether an action is intentional under a description is likely to be a function of features of the linguistic context whereas the conditions to be offered below are based on the causal structure of the event vis-à-vis solving the Many-Many Problem.
[10] Other notions of automaticity and control will be of use in cognitive science. I suggest that in each case, theorists invoke the simple connection and then explicate control as a type of top-down modulation. For agentive control, the type of top-down modulation involves solving the Many-Many Problem.

issues regarding how to construct such hierarchies, but since the notion is widely deployed

in cognitive science, I will assume it. The specific assumption is fairly minimal, namely

that intentions are closer to the top of any hierarchy than the basic perceptual, motor, and

memory processing that solves the Many-Many Problem.

Here, then, are the definitions:

(AC) Agentive or Intentional Control

For any subject S's token behavior of type Φ at some time t:
 S's Φ-ing is *agentively controlled* in respect of Φ iff S's Φ-ing is S's execution of a
    solution to the appropriate Many-Many Problem given S's intention to Φ.

When there is control, S's Φ-ing is understood to be an action-event, a concrete particular

whose causal structure is as described in the right hand side of the biconditional, and that

expresses the account of intentional action given above. The *appropriate* Many-Many

Problem is one such that the solving of it suffices for that event to be classified as a Φ. For

example, where Φ is kicking a football, S's kicking of a football is agentively controlled in

respect of Φ because S's kicking of a football is S's solution to the Many-Many Problem

given S's intention to kick a football. In short, we control what we intend.

We can define (agentive) automaticity as follows:

(AA) Agentive Automaticity

For any subject S's token behavior of type Φ at some time t:
 S's Φ-ing is *agentively automatic* in respect of Φ iff it is not the case that S's Φ-ing is S's
    execution of a solution to the Many-Many Problem given S's intention to Φ.

This is just the simple connection: the automaticity of Φ in S's Φ–ing is equivalent to the

absence of control of Φ. On this account, most of the Φ-s S can be said to do at a time will

be automatic. In our example, S's kicking with a force N, S's moving his foot with

trajectory T, with acceleration A, at time *t,* and so on. Finally, we can also define a strong

form of automaticity, what I shall call *passivity*.

(P) Passivity

For any subject S and token behavior of type B at some time t:
 S's B-ing is *passive* if for all Φ under which S's B-ing falls, S-'s Φ-ing is agentively
    automatic.

Where an action exhibits only automaticity, then the agent is passive and there is no

intentional action or agentive control. Let us now put these notions to work.

6. *Resolving The Threat from Automaticity*

Paradoxically, automaticity seems to imply the absence of agentive control and yet is a

critical part of skilled agency. We can now eliminate the air of paradox. On the one hand,

the definition of automaticity relies on the simple connection and automaticity as passivity

does imply the absence of agency. On the other hand, a process can be both automatic

and controlled with respect to different types of activities that the agent exemplifies at any

given time. That an event falls under an action type with respect to which it is automatic

does not imply that it is not an intentional action, so long as it does exemplify control as

per AC for some Φ. What is disallowed is that an action can be simultaneously automatic

and controlled with respect to the same Φ at the same time.

Consider playing the piano again. In aiming for automaticity, we relinquish some

control but not all. We control the type of thing that we intend to do, say playing a certain

work. We thus act intentionally. In playing the piano, the automaticity aimed for is that the

specific notes played need not be represented in one's intention. "Parameter specification"

is automatic because no top-down modulation at the level of intention is required to

specify the specific notes played, the ordering of fingering and so forth. Certainly, in

18

learning that passage, one can act with a changing set of demonstrative intentions, say to play *that* note with *these* fingers, and this is attentionally demanding. One has to attentively focus on relevant notes or relevant keys. But once the piece is mastered, setting those parameters is automatic. Playing the piece involves both automaticity and control, and the simple connection, properly understood, shows us how this can be true.

Actions as events are an agent's exertion of control *simpliciter,* as we might say. We secure the presence of action by ascertaining that the corresponding path in behavioral space is a result of an intention-guided solution to the Many-Many Problem. The focus thus far has been on successful intentional actions, and where the Many-Many Problem is successfully solved, it follows that there is an action event (agentive control *simpliciter*) and also agentive control in respect of the intended action type as per AC. One form of action worth considering at this juncture is unsuccessful action. While this is defective action, it is still action. Its status as action is secured in light of an underlying, aborted solution to the Many-Many Problem where what selections were made conform to the relevant intention. The question is whether there is, in all such cases, something that the agent controls as per AC? The thought above is that agentive control *simpliciter* (the existence of an action event) implies that there is agentive control with respect to some Φ as per AC. Yet there may not be any Φ subject to control.[11]

Let me raise two points as gestures to a fuller response. First, the content of our intentions are typically more complex than their expressions in language, say that one is going to Φ. That is, the content typically includes the more specific ways one will Φ, though never to the fine detail of parameter specification. Thus, one not only intends to make tea, but to make tea with this mug by pouring from this pot and so forth. So, there are

---

[11] Here, I respond to queries raised by Julian Kiverstein.

many opportunities for control, as per AC, even in aborted processes as when one fails to make tea because there the mug cracks from the heat. The pouring from the pot was under the agent's control because that Φ is in the scope of the intention. Second, as Falvey has noted, there are correct descriptions of action that appeal to the openness of the progressive. One can be said to be Φ-ing while one is doing just that, before the action is completed. So, I am making tea as I intend even if, when the mug breaks, the action is a failure. When such descriptions of the process are appropriate and reflect what we intend, we can be said to be exerting control in such cases.

AC allows for gradations of control. For example, over time, one's playing of an instrument might show increased automaticity over a defined set of features and thus less control in respect of those features. Yet that suggests that acquisition of skill means loss of control.[12] My piano instructor imparted a useful insight that is relevant here, namely that one practices not to play a piece in a specific way, say to imitate a performance of a Brahms intermezzo by Radu Lupu, but rather to be able to play in whatever way one wants. The control found in skillful action is that by automatizing a host of basic features, the skilled agent opens up behavioral possibilities that were not available before.

The putative threats mentioned in the introduction are threats to the will and thus threats to any of its forms such as free will or conscious will. It is important to see that the threat thus aims at agency itself and that avoiding the threat requires incorporating the notion of automaticity into the theory of agency. I have defused the threat of automaticity by showing that while automaticity is defined as the absence of control, its presence does not imply the absence of agency.

---

[12] I owe this observation to Till Vierkant and the way to respond to it to Jack Kurutz.

7. Settling Disagreements

Psychologists gave up the simple connection and, consequently the possibility of dividing kinds of processes between the automatic and the controlled. Reinstating the simple connection allows us to individuate processes along the automaticity/control divide, namely between actions and non-actions (passivity). Still, the metaphysical division between actions and non-actions is not the primary target of psychological research. Certainly, subjects in psychological experiments often perform actions (tasks) where the experiments investigate certain features of those actions that are characteristic of automaticity: insensitivity to cognitive load, parallel processing, bypassing of awareness etc. Of course, psychologists have also connected such features to broader questions about agency, and then they enter a traditional philosophical domain. I shall close with these issues in the last section, but I first revisit the puzzling disagreement among philosophers as to what counts as a mental action.

Let me begin with a disagreement with Strawson, namely his verdict regarding imagination and deliberation. On imagination, I noted that Strawson gave the correct answer: "given that one's imagining duly fits the specification [as intended] one may say that it is intentionally produced." While higher-ordered properties of action, say freedom or rationality, might require meeting further conditions, it is a mistake to look for action in any further sense than found in Strawson's characterization of intentional control, one that comports with AC. One's entertaining of an image of a pink elephant given one's intention to imagine that type of image is a reflection of one's agentive control. The intention explains why that specific path in behavioral space is selected. There are, of course, a host of automatic features associated with such imagination, say one's imagining the elephant as standing on one foot or facing left from one's point of view. One may not have intended

21

to imagine those elements yet that is what the mind automatically produced. By acknowledging these points, however, we do not reject that imagining as one intends is a matter of one's intentional control and thus is an action.

What of deliberation, where the "play of pros and cons is automatic"? The automaticity of specific thought contents as one deliberates does not imply that the process of deliberation is not an action, namely the solving of the Many-Many Problem in light of intending to determine whether *p* is true (theoretical deliberation) or whether to bring it about that *p* (practical deliberation). One's intention to do just that is a form of stage-setting in Strawson's sense that allows executing appropriate solutions to the Many-Many Problem. Accordingly, as the agent's intention plays a causal role in solving the Many-Many Problem, we have an action that is properly described as deliberating whether *p* or whether to bring *p* about. That the process can be categorized in terms of various forms of automaticity is not incompatible with its instantiating agency.

There are related phenomena where I might agree with Strawson, namely regarding the natural endpoints of successful theoretical and practical deliberation: judgement (belief) and decision (intention) respectively. These are, however, difficult matters, so for present purposes, I will assume that it is not possible to intend to judge (believe) that p or to intend to decide (intend) to Φ. *If* that is so, then that one judges that p or intends to Φ is by our account not something that the agent controls. Nevertheless, we should not lose track of these states or events as marking the *completion* of a process that the agent does control, namely deliberation. In focusing on one specific part of the process, namely its completion, it may seem that we do not have action at that specific point. There is nothing wrong in that thought, so long as we don't lose the forest for the trees, namely the endpoint as the culmination of solving the Many-Many Problem. Judging and deciding are just the

automatic culmination of an extended action. Indeed, the automaticity of judgments and decisions is perhaps what we should expect. Deliberators who aim to settle the truth or determine what should be done subject themselves to norms of rationality. Yet in properly subjecting oneself, there is nothing further to *do* once one recognizes the force of reasons as supporting a judgment or decision. One does nothing more than judge or decide. That is to say, one *automatically* draws the right conclusion.[13]

Finally, one can agree with both Fodor and Strawson (for the most part). With Fodor, we can acknowledge that we have all the leeway in the world in thought in that the extent of the Many-Many manifold that constitutes the Many-Many problem—the space of behavioral possibilities—is very large. In thought, is as large as what we remember and can bring to awareness for various tasks at a given time. Yet, with Strawson, we can acknowledge that there is also a way in which action is just a reflex or, better, automatic, in that many of the features of what we do are not subject to our control though some features must be.

8.   *Agency, Automaticity and Control*

I want to close with some reflections on the picture of agency that arises from the previous points. Actions, mental and bodily, are specific ways of solving the Many-Many Problem, specifically solutions that are constrained in the right way by one's intentions. Where there are such processes, we have agentive control. The crucial feature of this perspective is that

---

[13] These preliminary thoughts reflect an exchange with Matt Nudds. I noted earlier Mele's claim that remembering that *p* is never an action. Mele claims that the only action is bringing it about that one remembers that *p*. That view might be consistent with my account. One cannot, it seems, intend to remember that *p*, since one has thereby remembered that *p*, as *p* is in the content of the intention. Moreover, if bringing it about that one remembers that *p* (automatically, then) is a way of describing the process of solving the Many-Many Problem and not of just stage setting, then Mele and I are in agreement. I would add, however, that one can intend to remember *this* person's name and that can be subject to control even if one can't intend to remember *this* person's name as "John" since that would constitute remembering the name.

the mark of agency is internal to an event. Its status as an action does not reside in its antecedent causes but in its internal structure, the intention-constrained selection of a path in behavioral space at a given time.

There is a lingering worry given my acknowledgement above that intentions might always be automatically generated. How can action count as agentive control when its source is a matter of automaticity? It is a compelling thought that the agent must make herself felt at precisely such points when control threatens to evaporate, to reassert control so as to stave off its loss. The point of the current perspective is that control just is the role of intention in structuring a solution to the Many-Many Problem, full stop. The question of how the intention arises is an important one, but not one about agency in the basic sense. Rather, it pertains to whether the resulting action has higher-ordered properties such as whether it is free, rational or moral. I do not deal with these questions here, but simply emphasize two different questions: one about the conditions for agency; the other about conditions for its different forms.[14]

Given the Many-Many Problem, the pervasive automaticity of agency is what we should expect: we cannot intend and thereby control all the various ways of solving the Many-Many Problem even once a specific path is selected in intention. We may intend to grab an object or to recall a specific image, but the specific parameters that must be filled in to instantiate the solution to the Problem are not things that we explicitly intend to bring about, and thankfully so. It is not automaticity that is unbearable. What would be unbearable is its absence.

Psychologists have shown that much of our behavior is automatic, yet they have also spun this result as threatening to agency, leading to a discomfiting picture of human

---

[14] Andy Clark raised various issues to which this paragraph is meant as an initial reply.

beings as zombies and automatons.[15] One recent line of thought is the lack of conscious control in action. As Bargh and Chartrand put it, our "conscious intentions" are generally not in control of our behavior. Talk of *conscious* intention is common enough in the empirical literature on agency, but I must confess to not being certain what cognitive scientists are referring to. Do they mean an intention made conscious, namely an occurrent state where the content of the intention is at the focus of my awareness? A thought about my intention, namely a second-order occurrent state where my first-order state of intention is at the focus of my awareness? In normal action, I typically find no such thing. Rather, intentions are persisting non-phenomenal mental states of subjects that coordinate and constrain one's meandering through behavioral space. As I type these words, I am executing my intention to finish the essay by the deadline in a way consistent with my other intentions that are also operative (e.g. that I need to get to a meeting at noon), but there are no correlated conscious intentions or related meta-cognitive forms of awareness. Thank goodness! That would be distracting. Of course, I can bring the content of intention to consciousness on reflection, but that is a special case where I re-evaluate action. In general, when I act, my intentions are preparations to respond to the environment in certain ways, or as Strawson says, stage setting. Agentive control does not require that the intentions be conscious in either of the senses noted above.[16]

That our actions often bypass this form of conscious control is no sign that we are somehow not at the reins in our actions. We are at the reins to the extent that what we do

---

[15] Bargh's work, in particular, on the automatic processing of stereotypes and their influence on behavior yields surprising and interesting results.

[16] There are *decisions* and these may be conscious events (though they can be unconscious as well). Perhaps this is what Bargh and Chartrand mean when they speak of "deliberative choices". Still, these do not accompany every action, though when they do, they may contribute to action by setting one's intentions. There are also questions about unconscious vision that is also behind talk of zombie agents. On this, see my (under review) which is a response to Mole (2009). Briefly, while I do believe some visual representations guiding action are unconscious, I do not think there is sufficient empirical evidence for any stronger claim.

is a result of intentional control in the basic sense that our behavior is the intention-guided

solving of the Many-Many Problem. There is no denying that we are often moved to act,

and on the antecedents of action rest important questions about the rationality, morality

and freedom of our actions. But all these are higher-ordered properties of agency. Agency

itself is an internal feature of certain processes, our navigation through a complex world

that throws at us Many-Many Problems.[17]

References

Anscombe, G.E.M. 1957. *Intention*. Oxford: Blackwell Publishers.

Bargh, J. A. 1994. "The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition." *Handbook of social cognition: Basic processes* 1: 1–40.

Bargh, J.A., and T.L. Chartrand. 1999. "The Unbearable Automaticity of Being." *American Psychologist* 54 (7): 462-479.

Bargh, J.A., and M.J. Ferguson. 2000. "Beyond Behaviourism: On the Automaticity of Higher Mental Processes." *Psychological Bulletin* 126 (6): 925-945.

Fodor, Jerry A. 1983. *The modularity of mind: an essay on faculty psychology*. MIT Press.

Hommel, B. 2000. The Prepared Reflex: Automaticity and Control in Stimulus-Response Translation. In *Control of Congitive Processes: Attention and Performance XVIII*, 247-273. Cambridge, MA: MIT Press.

James, William. 1890. *The Principles of Psychology, Volume 1*. Boston, MA: Henry Holt and Co.

Koch, C, and F Crick. 2001. "The zombie within." *Nature* 411 (6840) (June 21): 893.

Logan, GD. 1988. "Toward an instance theory of automatization." *Psychological Review* 95 (4): 492-527.

Mele, Alfred. 2009. Mental Action: A Case Study. In *Mental Actions*, 17-37. Oxford University Press.

Mole, Christopher. 2009. "Illusions, Demonstratives and the Zombie Action Hypothesis," *Mind,* 118, (472,): 995-1011,.

O'Brien, Lucy, and Matthew Soteriou. 2009. *Mental actions*. Oxford University Press US, September 1.

Palmeri, Thomas. 2002. Automaticity. *Encyclopedia of Cognitive Science*. London: Nature Publishing Group.

Peacocke, C. 1998. Conscious Attitudes, Attention, and Self-Knowledge. In *Knowing Our Own Minds*, 63-98. Oxford: Oxford University Press.

Schneider, W., and R.M. Shiffrin. 1977. "Controlled and Automatic Human Information Processing: I. Detection, Search and Attention." *Psychological Review* 84 (1): 1-66.

Schneider, Walter. 2001. Automaticity. *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.

Shiffrin, Richard. 1988. "Attention." *Stevens Handbook of Experimental Psychology*: 739-779.

Wu, Wayne. 2011. "Confronting Many-Many Problems: Attention and Agentive Control." *Noûs* 45 (1) (March): 50-76.