T. Nathan Mundhenk, Michael A. Arbib

1 Introduction

An interesting question in research involving action and perception is *how do we as humans know the agency of our action* that is, how do I discern my actions from that of another person? In this, when I move my hand, I know I moved it, I also know that someone else did not move it. The same goes for speech, when I talk or have a verbal though I can reliably know that I in fact spoke or that a verbal though I had was a thought and not the utterance of another person. While such a distinction may seem trivial, it is brought into attention because of specific disorders where agency seems to break down. For instance, as will be explored in greater detail later on, people with schizophrenia commit actions such as verbal utterances and movements that they erroneously attribute to other agents. Additionally, they erroneously attribute the actions of others to themselves.

To understand how agency and schizophrenia tie in it is important to investigate what is known about schizophrenia. Observable symptoms, which are used to diagnose it, include delusions, hallucinations, disorganized speech, disorganized or catatonic behavior, and what are typically referred to as negative symptoms (affective flattening, alogia, or avolition). Hallucinations can take on the form of a constant verbal running commentary or even include two or more voices conversing with each other (DSM-IV, 1994). In schizophrenia, the auditory verbal hallucinations (AVH) and movements of the hand that are attributed to other agents (delusions of influence) are of particular interest to us here since AVH seem to be generated in a manner very similar to regular speech production (Stephane, Barton, Boutros, 2001) and are almost observable in the form of sublingual vocalizations. This then suggests that the voices that are generated by AVH as well as delusional hand movements are created by the agent, but for some reason, the agent does not have proper knowledge of the source agent. In other words, a schizophrenic patient generates an action, but does not attribute the generation of that action to himself.

1.1 The agent in us

When I execute an action I have an idea that I am going to commit this action, otherwise, I would be surprised when I received feedback on that action. Further, it seems important for the idea of learning that I

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

must store some intention of what I wanted to do (Suri, Bargus, Arbib, 2000)[1]. This idea theoretically can be stored in many ways. For instance, if I utter a sentence there are several things in what I said that I might want feedback on. For instance, I certainly would want feedback on the phonetic structure of what I'm saying. In this way I learn to move my mouth correctly to create the proper sounds that make the words I utter. I might also want feedback on the form of my sentence, how did the meaning of what I said sound. Additionally, I might also want feedback on the effect what I said had on the person I was speaking to. Did it seem like they reacted the way I expected. In order to do these things I need a memory of the production to compare against, since without it, it becomes very difficult to create comparisons. Thus, what did I expect my words to sound like, what meaning did I expect to convey, what reaction did I expect. I cannot expect to gain an understanding of the error from my action unless I have a baseline to compare it to. Further, I might also want memory to play a role online during the creation of action (Fletcher, Shallice, Dolan, 2000). For instance, if I have a memory of an intention of what I want to say, I might need to hold that memory until I am at some level satisfied with the next level of production. That is, I might create different sentences or sentence fragments until I believe that what I have constructed is what I want to say. In that way, a working memory that may be created, which holds my intentions, serves a dual purpose. First it holds a copy of the intended action for feedback from the world, second, it serves as a blue print while the next more complete representation of my action is assembled. Here as we discus working memory, we will be talking about a top-level intention. That is to say, it is the initial idea of what I would like to do. It is where I would like to move my hand to, or it is the idea I would like to communicate. However, it is in a state prior to being assembled into a coherent action. This assembly is left to areas such as Broca's area for instance (Need Citation) while the working memory as we will talk about in much greater detail is believed to reside in dorsal later prefrontal cortex (dlPFC) (Williams, Rao, Goldman-Rakic, 2002; Perlstein, Dixit, Cohen, *et al*, 2003).

In addition to having an understanding of what I am doing, it is also important to know what other people are doing. For instance when Laurent says, "The server is thrashed save your files!" I must know that the file server is going down and I should save my work. This of course is an obvious component to everyday

---

[1] Need another Citation. Not sure if this is totally on topic

life. I say things, other people say things, and as Homo sapiens we manage to understand each other. Further, this applies to other person's actions as well. If someone is going to hit me it is good that I have an understanding of their arm and hand and thus know what they are going to do with it. Also, if someone polishes and apple, and I might just want to polish one myself, it is handy if I can observe how they do this. Thus, I now have both a notion of action, what's being done, and agency, who's doing it. Further as has been proposed, humans and many other animals have a way of placing themselves in the actions of others (Rizzolatti, Fadiga, Gallese, Fassi 1996; Arbib, Billard, Iacoboni, Oztop, 2000; Frith C, 2002). In this action-mirror paradigm, I imagine myself moving my hands in the way another person does, or I imagine saying something like another person. Additionally I do this while the other person is executing their actions. I, in essence, am placed mentally inside of them. In this, the line between action and agent becomes less distinct. For instance as will be disused in further detail later, when I talk I must know the agent is me, if I don't I might logically deduce that someone else is talking, not me. I as a human must have some way of knowing when I talk, the words coming into my ears belong to me. I must also know that when someone moves my arm, perhaps because they bump into me, it is not me who is creating the moving.

Additionally it is important to know the agent of my target action during an imagining of their action. For instance, as I activate motor centers involved in both action imagination, and action creation (Rizzolatti, Fadiga, Gallese, Fassi 1996; Arbib, Billard, Iacoboni, Oztop, 2000; Frith C, 2002), my brain must be able to sort out simulation from the actual creation of the action. Otherwise, I could conceivably activate additional centers and execute the action fully, or the feedback might be mistaken for a real action. Thus I must track several things during both action sequences of imagination and action creation. I must track the intention for the action. That is, do I intend to execute it as a thought or as an external action? I must also keep in mind the source. Thus, from the activation of the brain areas involved in imagining actions, when they feedback to me, must allow me to know that I was imagining someone else moving, or speaking. It then becomes necessary to keep track of several things. The first is the action I intended, the second is the agent actor and thirdly, since this exercise always is in reference to me, *I imagine how to say something*, *I move my arm etc.*, I am implicitly included as a second agent. Thus, the working memory which we are exploring holds memories for actions and agent with intentional forms like, *I think to myself, "hello"*, *Bob*

says hello to me or *I imagine Bob throwing the ball*. This is not to say that all working memory is egocentric in this manner. However, as will be highlighted later, the symptoms of schizophrenia point to the existence for a more egocentric working memory in dlPFC. This is to say in short, when I have a though about saying something or commit to a movement, I store the intention of the action in working memory, that is what I intend to do. I also store the agent the action it linked to. Thus, I have knowledge of the action, and I have knowledge of the relating agent, and from this knowledge I have understanding of what to expect from myself and from others. Further, as we will talk in greater detail about, the linking between the action and agent, as held in working memory, is of equal importance. That is, the agent and action are stored in working memory such that they are bound together.

In the following pages we will discuss the idea of working memory in greater detail. We will try to understand its parameters by exploring neurocognitive, neurophysiological and computational approaches to both schizophrenia and working memory. We will also discus the outcome of our own simulation of a schizophrenic working memory.

1.2 Neurocognitive perspective

One of the major hallmarks of schizophrenia is a breakdown in the knowledge of agency. This comes in many forms. Schizophrenic patients hallucinate voices that they attribute to external agents, they have delusions that other people are creating their actions for them and they also have delusions of influencing others to act (DSM IV, 1994). In these examples, the common problem is that from an action that is created, either by the schizophrenic patient or another person, the observed agent that created that action is frequently incorrect. The patient did not will someone else to act, nor were they willed to act by another person, but this is what they perceive.

Similarly, in addition to hallucinatory symptoms it has been recently observed that patients with schizophrenia have difficulty determining whether they spoke or though an utterance (Franck, Rouby, Georgieff, *et al*, 2000; Brébion, Gorman, Sharif, *et al*, 2002). In studies, schizophrenic patients were compared to control subjects and it was found that they had greater difficulty when asked to recall correctly

if a sentence was spoken or thought to themselves. Schizophrenic patients also have difficulties determining if a hand movement they are observing is their own or that of the experimenter in condition where the movement of their hand is displayed on a video screen and is altered in trajectory and temporal synchronization (Franck, Farrer, Jeannerod, et al, 2001). It has also been observed as mentioned that schizophrenic patients can project their intentions onto other agents (Sarfati, Hardy-Baylé, Widlöcher, *et al*, 1997). For instance, a schizophrenic patient has the experience of controlling another agent as they observe that other agents actions. From this it seems that schizophrenic patients have difficulty in not only understanding the natures of their actions, but also have difficulties in terms of knowing who is committing an action even if the patient is creating an action overtly, covertly, and additionally they have difficulty if they are observing how the an action is being executed by another agent.

This leads to an interesting observation, in a world where we as humans must track multiple agents and actions; schizophrenics make almost exclusively egocentric attribution errors. If there was only a general mechanism for tracking actions or agents that is not egocentric, then if it was malfunctioning, one would expect that schizophrenic patients would not only hear voices that talk to them, but would hallucinate people talking to other people, for instance, hallucinating that you think two people in a room with you are talking. Instead, delusions and hallucinations seem to involve the self. That is, the mechanism for determining agent and action that is of interest here seems to have one marker pointing towards *me*. This happens whether *I* make another person act, another person acts on *me*, or *I* act upon *myself*.

The mechanisms of the action involved in schizophrenia have been hypothesized to involve working memory (Posada, Franck, Georgieff, Jeannerod, 2001; Perlstein, Dixit, Cohen et al, 2003). The most prominent theory to surface for the role of working memory in schizophrenia is that it is involved in a failure to understand ones own actions in a feedback mechanism (Frith, Blakemore, Wolpert 2000**;** Frith C, 2002). Thus, a normal person plans an action to execute, then executes the action. The observation of the action is feedback and compared to an action held in working memory. In a schizophrenic patient, according to the hypothesis, the record of the action is lost. Thus, a normal person would have complete knowledge of their actions. However, for instance, if a schizophrenic patient were to move their hand, they

would receive a surprise when their hand moves. We extrapolate from this that auditory verbal hallucinations are accounted for by the observation that auditory pathways are active during hallucination (Stephane, Barton, Boutros, 2001) and produce a verbal process of some internal voice, but since no record is kept of the voice being created, it is treated as external. That is, an utterance is created and progresses through verbal creation pathways, and returns as a vocalization *observed*, only to be dismissed as external since no record of it being created has been kept. Schizophrenic patients by this theory then confabulate the agent, which created the unknown self-created action. The confabulation then takes on a strong identity. That is, the source of an auditory verbal hallucination is confabulated to the point where the exact identity of the voice is guessed, even if the fictitious agent is nowhere to be found, or doesn't even exist.

A major question arises from this model of schizophrenia, why would a loss of knowledge lead to projection of the self onto another person's actions. Certainly, it is easy to imagine that if I forget the action, which I am executing, it becomes important for me to reason where the source of the action is coming from. However, why would I attribute myself into another agent? A partial explanation comes from theory of mind where I might place myself into another observer. That is, while observing another agent commit an action, I imagine myself committing that action myself (Arbib, Billard, Iacoboni, Oztop, 2000; Rizzolatti, Fadiga, Gallese, Fogassi, 1996). This can be bolstered by recent research showing that observation of self-action and of the actions of others activates a Mirror Neuron System (Arbib, Billard, Iacoboni, Oztop, 2000; Rizzolatti, Fadiga, Gallese, Fogassi, 1996). This system seems to be not only responsible to mirroring the actions of others, but its activity is egocentric, that is, it seems to only activate when the other agents actions are important to me. For instance, a Macaque monkey that observes a researchers hand may not necessarily activate its mirror neurons. However, if the researcher is doing something interesting like reaching for a raison, the neurons are far more likely to activate. While this could seem to answer in part how I would hallucinate making another person move, particularly if the feedback system for action is linked to the same mirror neuron system for both observation and self action, it begs the question, *why do I confabulate myself as the agent of another persons action?* One would think that if I were observing another person move, if they violated expectation by moving differently, I would still infer that they committed the action. That is, why do I place myself as the actor when clearly, it is far more

logical to infer them as the actor. Thus, a new suggestion here is that instead of not knowing the agent because of never having stored it or because I forgot it, I have what I think is explicit memory of the agent. That is, for some reason, when observing your action, the working memory I have for the agent is *me* instead of *you*. Not only would this explain how I can project my actions onto you and you onto me, but it explains how AVH hallucinated agents seem to have such a clear identity to those experiencing the hallucinations.

It should be noted at this point that in order to misattribute an action that you create, an involuntary component may be necessary, but is not sufficient. This can be seen in Tourrete's disorder where people will commit involuntary utterances, tics and movements, but it has not been shown that they lack an understanding that they committed the action themselves (DSM IV, 1994). This can also be seen in various alien hand syndromes where patients report that their hand executes meaningful movements, but that they do not have knowledge of having created the actions. Still they know that they created these actions. Thus, execution of a movement for which I did not necessarily decide to commit, does not cause a lack of agency understanding, which could be for two reasons. The first is that knowledge of the action is available, but the action was able to leap out through a lack of neural suppression, that is, I did not intend to execute the action, but was unable to suppress it (**?**). The second is that, knowledge of the action may have in fact been lost, but the person is able to handle the exception. That is, they may not have explicit knowledge of the agent, but they make a very logical inference as to the fact that they themselves committed the action and are not fooled into thinking otherwise. It is interesting to note that while diseases such as Tourrete's disorder are treated by the similar means as schizophrenia, namely with neuroleptics (Merck Manual, 1992) and seems to involve involuntary actions, several differences should be highlighted. The first is that involuntary actions in Tourrete's disorder are fully observable (DSM IV, 1994) where as involuntary actions in schizophrenia are as mentioned more covert, the casual observer generally only knows if the action is odd if a schizophrenic patient claims it is. Further, there are several diseases, which cause forms of alien hand syndrome. In these disorders patents execute various types of hand and grasping movements which they do not recall executing. However, if asked, they will not attribute the action to an external agent. (Inzelberg, *et al* 2000, Carrilho, Caramalli, Nitrini, *et al*, 2000) Instead, they admit that what they

executed was a self-created involuntary action. Here again, we see that involuntary actions are not enough to create the type of agency disruption seen in schizophrenic patients.

Another characteristic of interest with regard to agency in schizophrenia is the observation that schizophrenic patients seem to latch onto themes during discourse (**Jorgensen, Jensen 1994; Jorgensen 1994**). That is, even though what a schizophrenic may say may seem very unorganized, they still seem to stick to a theme in their discourse. This is characterized as an almost obsessive trait in schizophrenia. Similarly, the hallucinated agent talking to a schizophrenic patient frequently follows a thematic course (**Jorgensen, Jensen 1994; Jorgensen 1994, DSM IV**). This can be highlighted as an underlying thought process when compared against results from the Wisconsin Card Sort Test (Butler, Jenkins, Sprock, Braff, 1992; Franke, Maier, Hain, Klinger 1992, Lanser, Berger, Zitman, *et al* 2002) on schizophrenics. In this test, cards are shown to subjects and they must learn the paradigm of the cards shown to them. This is in essence an open experiment in which an abstract theme must be learned by observing the cards with no highly overt signals from the experimenter. In the middle of the experiment the cards change paradigm and the subject must then figure out that the paradigm has shifted and what that new paradigm is. It has been shown repeatedly that schizophrenic patients persevere in their understanding of the underlying paradigm. In other words, when the paradigm changes, schizophrenic patients seem to have difficulty noticing the change and adapting to it. It's as if they can latch onto the first paradigm, but will not let go. We believe this underscores what may be an inability of schizophrenics to release a task memory. As we discuss in the modeling perspective section, models of working memory have shown that an active clearing of contextual memory is needed in order to prepare contextual working memory for a new paradigm (Rougier, O'Reilly, 2002).

Another important observation about symptomology and schizophrenia is that it seems to affect most strikingly language and grasping actions. For instance, hallucinations in schizophrenia rarely involve visual sensations (Merck Manual, 1992). It also seems from our review of the literature, that for movement type actions, grasping or hand movements are the kind most often hallucinated. Delusions of influence, in most cases, is where a schizophrenic hallucinates that another agent is causing their hand to move. While it is not

as common as auditory verbal hallucinations, it seems to be the prevalent form of delusional movement studied, much less found in the literature (**dudes**). Further, asymmetries have been observed in the grasp of schizophrenic patients (**Purdon, Woodward, Flor-Henry, 2001**) and most importantly is the difficulty schizophrenic patients have in determining whether a hand executing an action in a video screen, similar to their action is their hand or not (Franck, Farrer, Jeannerod 2001). While, such a studies do not necessarily imply a full-blown agency attribution problem, since the affects observed are most common in schizophrenic patients with delusional disorders such as delusions of influence, it is reasonable to assume that a similar cause may underlie both external agency misattribution and internal agency misattribution. That is, what causes me to misattribute my hand moving to me may be the same mechanism that causes me to misattribute my hand moving to you[2]. As such, the idea here is that both language and grasping share a similar pathway for agency attribution as well as for action and perception connection. This creates a theoretical connection to, as well as a reciprocal foundation from the theory that language evolved from grasping (Rizzolatti, Arbib, 1998). In this case, because the most profound positive symptoms of schizophrenia seem to involve language and grasping, if schizophrenia has a single underlying cause, it shows that these systems are closely linked. Further, since the fault in schizophrenia seems to lie in part with recognition, especially for positive symptoms such as auditory verbal hallucinations and delusions of influence, it lends to a theory that a single recognition system may be involved, in this case, for instance the mirror neuron system. While we don't directly implicate the mirror neuron system in the failures of schizophrenia in agency attribution, since the mirror neuron system is theorized as allowing us to place ourselves into others actions as well as observe our own actions, and since it is this activity of observing our actions and others actions that is involved in agency misattribution of schizophrenia, we believe that this shows evidence for a unified pathway for language and grasping at the level of agency and action through the mirror neuron system.

To explain how it seems that actions and agency can become disjoint, it has been proposed that some of the symptoms observed are the result of a temporal binding difficulty (Haggard, Martin, Taylor-Clarke, Jeannerod, Franke, 2003; Garcia-Toro, Blanco, Gonzalez, Salva, 2001). Such an idea began to surface

---

[2] I'm not sure if this is a bit contrived

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

when it was discovered that schizophrenic patients seem to have problems with associative memory tasks such as latent inhibition and kamen blocking (Serra, Jones, Toone, Gray, 2001; Bender, Müller, Oades, Sartory, 2001). Since association in learning seems to be product of temporal binding in many cases (Suri, Bargas, Arbib, 2000,?), we hypothesized that analogously schizophrenic patients may not be able to bind sensory experience with the proper agent in the same way. This is further illustrated in studies that show that loss of recognition of ones own actions in schizophrenia seems to increase with a larger temporal disjoint between the observed action and the action being executed by the patient (Haggard, Martin, Taylor-Clarke, Jeannerod, Franke, 2003). In this, it seems that schizophrenic symptoms may be explained in part by an over correlation between an action observed and the action held in working memory. That is, for instance as will be discussed later in greater detail, if a system of neurons, for instance dopamine neurons, which play an important part in temporal difference associations (Suri, Bargas, Arbib, 2000, Waelti, Dickinson, Shultz 2001), is up primed more often than it should be, then it may signal that on observed action is the same as an action held in memory more often then it should. This could be because it sends out more signal if it tends to stay in this up-regulated state. That is, in a healthy person, the more an action observed is like a memory of an action, the more neurons that look for correlation will fire. Surpassing a firing threshold may signal recognition. If these neurons are up regulated, they will fire more vigorously in general, and be more likely to fire above threshold when an action observed and one held in memory are only weakly correlated.

In recent research, it was found that agency attribution errors in schizophrenia may indeed be due to some sort of temporal binding problem (Haggard, Martin, Taylor-Clarke, Jeannerod, Frank, 2003). In the study, they showed that schizophrenic patients were more likely to perceive timing between two key presses on a keyboard as being closer together. Thus, schizophrenic patients observed that a second key press happened around 50 ms after the first when in fact it had happened 250 ms later. Their conclusion was that hyperpriming from dopamine neurons caused a stronger binding between the two events. Additionally we believe that a residual memory effect as we have described may also explain this since the memory for the subjective markings of time may be intermixed. That is, the time held in working memory for schizophrenic patients may be in a half way state between the first working memory held and the second

one. Since the bias in timing from their experiment drifts both ways (the observed time for first press moves towards the second and visa versa) it is easy to see how the memory representation may be half way between.

Another factor that should be mentioned at this point is the observation that schizophrenic patients experience hypofrontal activation in PFC during working memory tasks (Hazlett, Buchsbaum, Harvey, *et al*, 2000; Perlstein, Dixit, Cohen, *et al*, 2003). While a hypofrontal activation might indicate that a stronger holding of working memory is less tangible, we argue the opposite. This is because PFC is probably also involved in planning and dealing with memory exceptions (Kobayashi, Lauwereyns, Hikosaka, *et* al, 2002; **Leon, Shadlen, 1999**). [3]That is, if I have task that stresses working memory and PFC more, I would expect that it would work hard because I would have to spend more time inspecting, replacing and processing data in PFC. Such observations can be seen in studies that show that as task complexity and load increase for working memory tasks, PFC becomes more active (Hazlett, Buchsbaum, Harvey, *et al*, 2000; Perlstein, Dixit, Cohen, *et al*, 2003). If I am schizophrenic, and if I think I always know the answer, that is the dlPFC states in a working memory paradigm that it knows the answer more often than it should, it will work less hard to process working memory information in search of a new correct answer. Thus, because a schizophrenic patient may think they know the answer more often then they in fact do when challenged with a working memory task, you would expect to see hypoactivation since the additional working memory resources needed to complete the task are not requested. This could be analogous to brain imaging studies that show that people who are more practiced on working memory tasks begin to use fewer resources in dlPFC (Petersson, Elfgren, Ingvar, 1999). Thus, the schizophrenic brain thinks it is expert in a working memory tasks, but in fact it is not. This idea can be further expanded by research that shows that lateral PFC areas seem to be more likely to become active in a task where subjects areas asked to link words of greater and greater semantic difference (Fletcher, Shallice, Dolan, 2000). The authors concluded that lateral PFC areas are activated more by the challenge if linking together items of greater semantic difference then from an increased demand on working memory storage.

---

[3] get papers off Petersson Paper

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

In summary, from a neurocognitive standpoint we are theorizing several things (1) a working memory exists that stores agent and action for both language and movement (2) the mechanism of interest is egocentric (3) The memory of the agent as it relates to the action may be intact but is mismatched. (4) Positive symptoms of schizophrenia may be characterized by a temporal binding problem.

1.3 Neurophysiological perspective

(TABLE)[4]

To try and build such a case it is important to not only review literature of functional neuroimaging and cognitive symptoms, but one must also look at neurophysiology to understand how a theory can be realized. To start with, for schizophrenia, fMRI studies have concentrated for the most part, on abnormal volumes and activations in such areas as the hippocampus and temporal lobe (Thompson, Vidal, Rapoport, *et al*, 2001), amygdale (Nieman, Hammers, Coenen, et al, 2000), Basal Ganglia (Corson, Nopoulos, Miller, et al, 1999) and Dorsal Lateral Prefrontal Cortex (McDowell, Clementz, 2001). Additionally as we discus studies have also been conducted on the morphology of neurons primarily in the prefrontal cortex and hippocampus.

When studying neurophysiology, much evidence has been ambiguous. Most studies seem to concentrate on the role of 5-HT2a serotonin (Dean, 2003; Lewis, Kapur, Zipursky, *et al*, 1999; Eastwood, Burnet, Harrison, *et al*, 2001), DA1 dopamine (Koh, Bergson, Lidow, *et al*, 2003) and more recently parvalbumin GABAergic interneurons (Volk, Lewis, 2002; Beasley, Zhang, Reynolds, *et al* 2002). Further, we have observed that in terms of brain area, the greatest attention seems to be focused on dorsal lateral prefrontal cortex (dlPFC) and the hippocampus.

Studies of neuromorphology show degradation of local axonal parvalbumin GABAergic interneurons in dlPFC (Volk, Lewis, 2002; Beasley, Zhang, Reynolds, *et al* 2002). These neurons are fast spiking interneurons that suppress other neurons including themselves (Gao, Goldman-Rakic, 2003; Seamans, Gorelova, Durstewitz, Yang 2001; Tanaka 2000). Activation of these neurons comes primarily from

---

[4] Insert table of neurochem here

influence of pyramidal neurons. As will be discussed later, the pyramidal neurons that control these interneurons have also been implicated as having an abnormal morphology.

Also of particular focus are DA1 as well as DA2 dopaminergic neurons in PFC given that early treatments for schizophrenia were dopamine antagonists (Merck Manual, 1992). Also, well-known dopamine agonists such as amphetamines seem to exacerbate symptoms of schizophrenia (Merck Manual, 1992). Here there is some disagreement on whether abnormal morphology exists. For instance studies of the D1 receptor-interacting protein Calcyon show an abnormality in dopaminergic expression in PFC (Koh, Bergson, Lidow, *et al*, 2003), but levels of the dopamine metabolite homovanilic acid were found in very early studies to be normal thus suggesting the opposite that dopamine levels are normal (Bowers, 1974). It is thus to this day that the dopamine hypothesis while popular, remains controversial (**Kestler, Walker Vega, 2001**).

Interest has also been focused on serotonergic neurons, and in particular 5-HT2a. This has come about more recently with the development of the atypical antipsychotic medications used to treat schizophrenia. Such drugs, which seem to have fewer physiological side effects than earlier drugs are theorized to antagonize serotonin far more readily than dopamine (Merck Manual, 1992). However, here again abnormal expression seems to be uncertain (Lewis, Kapur, Zipursky, *et al*, 1999). Further, in studying the morphology of neurons in schizophrenic brains, complications seem to arise as a result of the drugs used to treat patients. It is thus difficult to tease apart the affect of the drug from disease specific abnormalities observed in the brain (Dean, 2003). However, it is important to note that increase in 5-HT2a serotonin expression seems to increase the ability of working memory fields to hold memories in dlPFC and increases their clarity (Williams, Rao, Goldman-Rakic, 2002). Thus, it is reasonable to conclude that an increase in serotonin if observed in schizophrenics would lead to an increased holding, or residual in working memory fields.

While these three neuron types have been the subject of many separate experiments, their connections to each other has also been of greater focus in recent years (Aghajanian, Marek, 2000; Seamans, Gorelova,

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

Durstewitz, Yang 2001; Krimer, Goldman-Rakic, 2001; Tanaka 2000). Thus, it seems that dopaminergic, serotonergic and GABAergic neurons interact with each other and exert reciprocal effects. This has been observed as suppression generated by GABAergic interneurons is modulated by dopamine neurons (Seamans, Gorelova, Durstewitz, Yang 2001; Krimer, Goldman-Rakic,), serotonin neurons (Aghajanian, Marek, 2000) and other GABAergic interneurons (Gao, Goldman-Rakic, 2003). These connections have been validated in the pre-frontal cortex (Gao, Goldman-Rakic, 2003; Krimer, Goldman-Rakic, 2001; Seamans, Gorelova, Durstewitz, Yang 2001; Aghajanian, Marek, 2000) thus, it would seem that in general, a neural model of schizophrenia would attempt to make use of this variety of connections. It also creates an interesting situation. Since all these neural types are reciprocally connected, if one of the neural types was diseased, one might hypothesize that its malfunction could spread to the other connected but distinct neural types. Thus, finding the root neural cause from an experimental standpoint may be confounded by the fact that all three neural types may become diseased, but only one is the source. The other neurons show abnormal morphology because their input messages are corrupted. This can also carry over into other brain areas as well. fMRI data on schizophrenia seems to be a veritable potpourri of possible diseased areas. For instance if the locus of the disease is the dlPFC, if it sends abnormal signals to other regions of the brain for instance through its connections to the hippocampus (Swanson, 1981) then the hippocampus could assume an abnormal morphology. Such a hypothesis could pan out since it has been shown that in schizophrenia dlPFC has a lower volume before disease onset, while the hippocampus has a lower volume following disease onset (Narr, van Erp, Toga, *et al*, 2002). Since schizophrenia is strongly hereditary, the dlPFC may have been set for failure since birth, with the hippocampus falling into a diseased state when the dlPFC finally does break down.[5]

From a neurophysiological standpoint, it is also important to look at the role of dopamine in other areas of the brain and how it relates to the dlPFC. Studies of dopaminergic medium spiny neurons in the striatum show that they play an important role in associative learning (Schultz, 2002; Waelti, Diskinson, Schultz, 2001; Suri, Bargus, Arbib 2000). Thus, it has been experimentally shown that these neurons can prime when they receive an input from an efferent connected neuron. Then, when a second signal comes in within

---

[5] I'm not sure if this sounds like a chicken our egg problem. It might be rather flimsy.

a certain time frame, the neuron fires vigorously. For these purposes, the neuron seems to play a role in linking two events together, in this case stimulus and reward. Its job in essence is that of a fast correlator for a hebbian style learning. Thus, within a short time span (100-300 ms) a dopamine neuron can be primed and ready for a second stimulus. As dopamine is added to these neurons, they attain greater ability to prime, that is, they prime much more readily. This takes us back to the argument for temporal difference in schizophrenia. If dopamine in hyper-primed in schizophrenia, and dopamine neurons act the same in dlPFC as they do in the striatum, then it is easy to imagine that actions observed could more easily correlate with actions stored in dlPFC working memory, thus creating a temporal difference binding error. However, since the dopamine hypothesis is controversial, over correlation from this type of temporal binding becomes less tangible. Again however, given the fact that the first drugs used to treat schizophrenia were dopamine antagonists (Merck Manual, 1992), its role cannot be discounted.

1.4 Computational modeling perspective

Additional insight into the mechanisms of working memory has been provided by a variety of computational neural models. For instance, a highly detailed model of PFC neural interactions was created in an attempt to shed light on the observed interaction of neurons in working memory fields (Derstawitz, Seamans, Sejnowski, 2000(1); Derstawitz, Seamans, Sejnowski, 2000(2); Seamans, Gorelova, Durstewitz, Yang 2001). In this model, memory fields comprised primarily of dopaminergic neurons are linked together and compete via their regulation of inhibitory GABAergic interneurons. A principle finding of the simulations conducted was that at high dopamine levels, neurons became resistant to outside input. Thus, they were able to filter out input from background activity in a functional environment. However, this also provides an interesting insight into schizophrenia since dopamine states in schizophrenia are hypothesized by many studies to be very high. From this one could conclude that if the simulation is valid, a high dopamine state could make a memory field more likely to hold a memory beyond its useful time. That is, an increased dopamine state could yield another temporal binding issue in the form of an increased holding or residual in working memory.

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

Another study of importance simulated the working memory in PFC responsible for context (Rougier, O'Reilly, 2002). In this model when a paradigm shift is needed, active working memory fields are cleared by a signal from an adaptive critic. Thus, during a Wisconsin card sorting task, when the task shifts to a new paradigm the context needs to shift with it. The simulations results apply to schizophrenia since as has been mentioned, schizophrenic patients have difficulty shifting paradigms during the Wisconsin card-sorting task. If the working memory model used for context is analogous to a working memory model for egocentric action recognition, then it also suggests that working memories, in the brain, during a working memory task, may need to be held until cleared during a context like switch. In this case, for a model of egocentric action recognition, a context switch might be expected outcome at the completion of the action.

In addition to these models, another model demonstrated the importance of the interaction between inhibitory interneurons and excitatory pyramidal neurons in PFC. In the experiment by Tanaka (2000) a memory field was constructed to agree with observations about the dlPFC working memory fields. Pyramidal neurons interacted by suppressing one another through interneurons that had either a preference for the same type of stimuli as the source neuron in that memory field, or a completely orthogonal preference. After a sufficient input from a stimulus, the network maintained the working memory signal. At the same time, other inputs were kept out. If interneurons were degraded, the network either lost its sharpness or had an increased activity. If observations about degradation in parvalbumin type GABAergic interneurons is correct, this study might suggest that decreased expression by interneurons would leave a working memory field more active and less sharp. This would lend itself again to the idea that working memories in schizophrenia are held onto, but are erroneous. That is, when an observation of an action is feedback, misattribution of the agent comes not from lacking a memory, but from having a corrupt memory. This is supported again by the neurological observations that GABA expression is decreased in schizophrenia while 5-HT2a serotonin expression and dopamine expression may be increased.

Other models that are important for this discussion include the extended TD model (Suri, Bargas, Arbib, 2000). While the models main thrust is to simulate dopamine based reinforcement learning in striatal neurons, it sheds light on how two events in the brain can be correlated even when they are temporally

separated. The model is applicable here because if dopamine neurons found in the dlPFC are very similar to striatal neurons, we would expect them to act the same. That is, we would expect them to fulfill a role of correlating two events with a temporal separation. In a model of the dlPFC what one might expect dopaminergic neurons to do is link the observed actions with the actions stored in working memory. In this, working memory neurons would prime the dopamine neurons, which would then become active when another signal is received via feedback that matches the action executed.

2   The Model

To investigate the role different neurons may play in agency we have constructed a model of working memory and recognition. The idea behind the model is to simulate working memory for the agent and the action being committed. In this case, we are at the tip of a feedforward model. A rudimentary intentional action code is created along with the corresponding agent that binds with that action, for instance *me*. These memories are held in working memory until the action is satisfactorily completed as observed in dlPFC working memory studies (Williams, Rao, Goldman-Rakic, 2002). This signals a feedback of a correct action completed.

The agent, which has been bound with the action, is the expected agent. For instance, when my hand moves and I made it move, when I observe it move in the way I intended, I assume I committed this action because I expected it to do so. Here such an assumption is carried out by correlating the action observed with the action expected held in working memory. When such a correlation happens then the agent in working memory is correlated to the observed action from its bonding to it. While much of the literature reviewed on the subject of working memory in dlPFC has focused on memory fields for intention on hand movements, namely trajectories, since we found no other studies about working memory for what might be described as word trajectories, we assume that the mechanisms for working memories in creating a meaning for language, are similar to those working memory fields that create trajectories for movement, which is another type of meaning[6]. In addition, we have added another working memory to address the

---

[6] I hope this isn't to weird a thing to say

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

agent for whom the action is directed at. As mentioned earlier, this agency working memory is egocentric in that, if more then one agent is considered in the memory, one of the agents always has to be the self.

In our model, it is our hypothesis, that given the evidence so far, an abnormal holding in working memory, that is, a strong residual memory in either the action or agent working memory fields causes delusions about actions and the agents related to them. This then creates a corresponding hallucination about an action. This is also a type of temporal binding mismatch since as we will highlight, the agent or action in the first epoch can be left over and bind to an agent or action in the next epoch. We also consider the idea that delusions and another type abnormal temporal binding caused by over activation of dopaminergic systems may cause hallucinations as suggested (Haggard, Martin, Taylor-Clarke, Jeannerod, Franck, 2003). The idea here is that when I commit an action, or have a thought about an agent that I believe will be connected to the action, when that action is complete, the agent in working memory should be stored to a longer-term memory then removed from working memory and treated as an expectation fulfilled. Instead, in schizophrenia, it holds on. When I commit to a new action the last working memory should be gone. Instead in schizophrenia, the last epochs agent is still held to some degree in working memory or as we will also explore, an over priming of dopamine correlational centers may cause the ability for the wrong agent and action to bind. Thus, when feedback for the action comes back, instead of attributing the action to the correct agent, as I should, if my brain is in a schizophrenic state, I attribute the action to another agent.

It is important to here to specify by what is meant by an action and agent. In the working memory we are proposing, one agent is always me. However, there is a second agent which is the agent acting towards me. Thus, if I have a thought, it is *me* thinking to *myself*. If someone talks to me, it is, *someone* talking to *me*. This is as mentioned because delusions and hallucinations in schizophrenia seem to have a predominantly self-centered mode. That is, voices talk to me, I control you, you control me etc. It does not come in the form of you talking to him. The fault in agency misattribution comes out of the second agent. The working memory seen in healthy individuals is of the form *I utter to me*. However, in a schizophrenic working memory, the second agent memory becomes corrupted and takes on the form *someone else utters to me*. Thus, auditory verbal hallucinations in the model are experienced when an overt and an imagined utterance
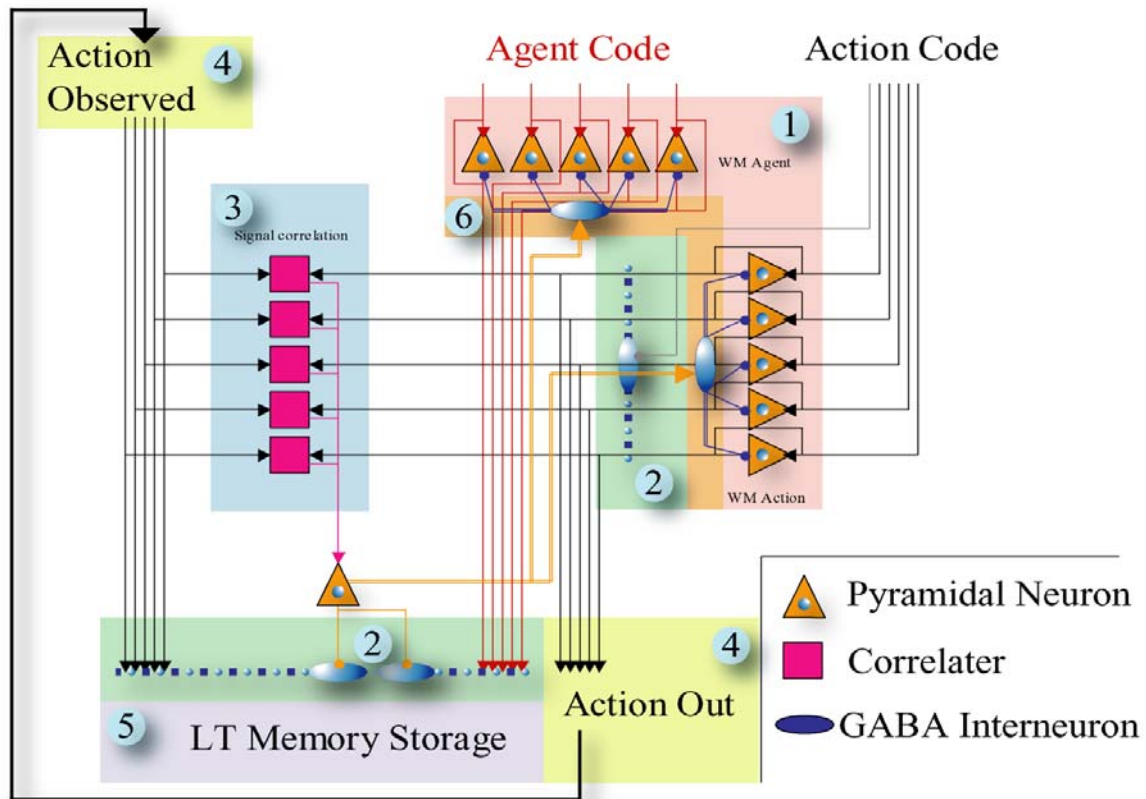
Figure 1.This is a basic model of the working memory that holds patterns of the agent and action codes in sets of pyramidal neurons. The network can be broken into 6 major sub-components. (1) W 5-HT2a based working memory field that holds initial action and agent intentions; (2) Gating neurons that allow actions to pass or allow working memories to be stored; (3) A GABA/DA1 correlator that associates incoming signals to working memories (see figure 2); (4) An action feedback loop that simulates an action output that is feedback to the system after delay; (5) a longer term memory that stores associated succesfult action/agent combinations; (6) Chandelier type parvalbumin neurons that wipe memory fields to allow new memories to enter.

are ambiguated from a residual memory as one follows from the other in a temporal misbinding. Further,

the agent is also ambiguated in the same manner. If I am schizophrenic and I have a thought about agent *A*

speaking, then I speak out loud, the action of thought and speech are ambiguated into a sublingual

vocalization, further the agent is also ambiguated, meaning that it could become *A* instead of me. That is,

the form of the outgoing intended action, which should be *I think to myself, Bob says hello to me* can

become intermediately meshed with a second action *I say hello to Bob* to form a new corrupt observed

working memory of *(I or Bob) (thinks or says) hello to me*. Sorting this out, you might then get one

combination of *Bob thinks hello to me* (Bob is controlling my thoughts) or *Bob says hello to me* (Bob is

talking to me when in fact he isn't) when the current working memory that was intended was *I say hello to Bob*.

There are several reasons for us choosing the idea of a residual memory for schizophrenia. First, it helps explain external delusions of influence. That is, it helps explain the phenomenon of schizophrenic patients having the delusion of influencing other people's actions. In a theory of mind type framework, I must have some sort of memory of their action, then I must attribute it to myself. It is hard to see why, in the absence of an agent or action memory, I would attribute myself onto another person's actions. It is easier to imagine how such an action might occur if I actual felt that I know that I was the agent. Secondly, it seems to agree better with neurochemical and neurophysiological data. Since schizophrenia is characterized most likely, by increased activation of dopamine, serotonin neurons or a decrease in GABAergic neurons, additional activation of memory fields seems more likely. Thirdly, it agrees better with observations about perseveration in memory tasks with schizophrenia. Fourthly, it explains the seeming determination that schizophrenic patients seem to have about the agent in their delusions.

It should also be emphasized at this point that we believe that working memory for language and action could overlap, but must maintain a mutual exclusivity. That is, the action code for a grasp could be held in the same working memory field as an action code for speaking. However, we make no assertion that this is or is not the case. We do however assert that the mechanisms underling both memory processes are similar enough such that they can both malfunction in similar ways as is observed in schizophrenia.

2.1 Model Components

The model is broken into several sub components each of which we will describe in greater detail later. Each sub-component is created in a manner that is as simplistic yet biologically feasible as possible. If we were uncertain that a certain type of neural connection was possible it was left out. Additionally as will be discussed, many of the components of our model bare similarities to other computational models of working memory, but since the goal of our study was different from theirs, namely the study of schizophrenia, different types of connections were emphasized.

The model components, which we will represent, are (1) a simple 5-HT2a working memory field (2) a GABAergic gating mechanism (3) a dopamine/GABA cooperative action/agent correlator (4) a feedback loop (5) a memory matching mechanism. (6) GABAergic memory field control. Figure 1 shows how these components are laid out.

2.1.1 5-HT2a Working memory field

The first part of the model, the working memory field, is a self-recurrent set of neurons designed to hold a memory for a short period of time. The memory held here we believe is an intention or expectation about an action or agent. It has not yet been coalesced into a sequenced action or utterance. It is merely the intention of what I would like to do. This memory is then held until either the neurons run out of steam, or an input from an inhibitory interneuron squelches the signal.

Models of working memory at this stage typically employ direct GABAergic inhibition between dlPFC memory field elements. (Durstewitz, Seamans, Sejnowski, 2000(1), Tanaka, 2000) However, the purpose for this in the model by Tanaka is to sharpen or hone a working memory for action. Since the working memory we have employed may be used for language, it was difficult to see how we might employ such an interneuron scheme. Additionally, the model by Durstewitz, employs interneuron connection between dopamine elements. We use a similar connection type in our dopamine-based neurons that correlate internal and external actions. Thus we hypothesize that activity observed from dopamine neurons in working memory could be a affected of by inputs from 5-HT2a serotonin neurons, also observed to activate in dlPFC working memory (Williams Rao, Goldman-Rakic, 2002), may be responsible for actually holding the memory while the dopaminergic system may be more responsible for its regulation via connections to GABAergic interneurons.

2.1.2 GABAergic gating mechanism

These are interneurons whose job is to gate a working memory from either storage in a longer-term memory or keep an action from being executed. Gating neurons are will known to exist in the basil ganglia

(**dudes**) **frontal eye field (dudes), cerebellum (dudes)[7].** It is also logical that they would play a part in working memory since it has been shown that a working memory of various actions can be held for reasonable periods of time, in dlPFC and other brain regions, before they are executed (Williams, Rao, Goldman-Rakic, 2002; Rizzolatti, Fadiga, Gallese, Fogassi, 1996). In our model there are two major gates. The first gate opens to allow an action held in working memory to be sent to other brain areas such as Broca's area to be processed into an action (overt or covert). The other gate allows the agent to be stored when the action it is bound to is observed. The purpose for this is that is allows an expected agent to be assumed as the one who executed the action without the need to wait for an agent recognition as well. That is, since you have a record of the expected agent held in working memory, it is much quicker to simply allow it to be stored as the agent that committed and action when that action was observed than to spend time making a full action/agent recognition. (**Get a citation**). This gating mechanism is also similar to one proposed by Grace (2000) who described a method in which a reinforcement signal may be sent onto thalamocortical areas. The primary difference is that we allow memory fields in working memory to control the gate instead of hippocampus. We also modeled GABAergic based gates for our model while he did not. However, it is important to note that the exact mechanism of our gate is less important than its function. Thus, we do not debate at this time the exact nature of the gating neurons.

2.1.3 Dopamine/GABA correlator

The job of this module is to compare incoming actions perceived with the one already held in working memory. Seen in figure 3, it functions as a fuzzy "and" circuit in basic terms. When a single signal comes into one side of the module, it activates a GABAergic interneuron, which inhibits through put of both this pathway and of the other pathway. Thus, if the two inputs are a working memory for action and another for action perceived, if only the action perceived enters, then it blocks itself off. The circuit gets its "and" ability since GABAergic interneurons also inhibit one another. Thus, when a signal is received to both action observed and action stored working memory, the GABAergic interneurons mutually inhibit each other, thus opening the gate for both inputs. The more identical the inputs are, the more the GABAergic interneurons will mutually inhibit and the more signal is allowed to pass. On the other hand the more

---

[7] I need to verify this information

disjoint a signal is, the more suppression the circuit exhibits. The power to inhibit is further amplified by having the two inputs received by a dopamine neuron that primes for an input. If some disjoint signal leaks through, its power cannot be magnified unless it has a reciprocal connection that is also active. In other words, if some of the action perceived signal is received by the dopamine neuron, it gets no benefit. However, if both perceived action and action working memory are received, then the dopamine neuron primes and creates a fast hebbian stile plasticity boost.

It is also important to note the importance of correlation in the PFC. As has been suggested; lateral PFC plays an important role in the semantic linking of words. From this we can hypothesize that the PFC has mechanisms built in for linking the meaning between semantic representations (Fletcher, Shallice, Dolan, 2000). Since our intentions as we have been calling them are roughly analogous to semantics, we can see that such a mechanism for linking actions intended to actions observed is highly feasible.

2.1.4 feedback mechanism

The feedback mechanism is very simple in our design. It takes the output action from the action working memory after the action has been gated through. It then waits for a simulated interval, in this instance a few seconds then repeats the action code into the dopamine/GABA correlator mechanism. It is done in this way since other than a temporal offset we have no reason to assume that the feedback code should be very different from the initial action code in working memory. Thus, we assume that feedback processing in essence reverses the process of the feedforward mechanism. It should be noted that we did not want to assume one-to-one transformation, so the feedback is an approximation of the feedforward. That is, the firing rate for feedback is made inexact by ramping it slightly. Thus, when feedback meets the feedforward command for action stored in working memory, the match is not exact. Since the brain is noisy (**dudes**) we should expect that feedback and feedforward signals should only resemble each other and not be exact duplicates.

2.1.5 Memory Matching Mechanism

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

In order to validate the choices the model makes with respect to which agent and action it matches up we have created a simple memory unit. In it we have placed a code that represents the agents and actions that will be executed as well as a few other extra control memories. When an action or agent code enters the memory matching it is compared against the memories stored using dopamine/GABA correlators like the ones described. The memory pattern that fires most similarly to the input pattern is the memory chosen. For instance, when action A is received, it is simultaneously compared with memories A,B,C,D and E. If the pairing of memory A with action A produces the strongest activity, as it should, then the decision is that action A was committed. This tests the existence of psychotic thinking in the network. If for instance we know we coded action A to execute, and the memory for action B had the highest response, then we know the computer made the mistake of observing that action B was executed when in fact action A was. Further, by observing the executions of two actions consecutively we cannot only observe that it mistakes an error for a given action or agent, we can also observe if the error is a perseveration error or a confusion error. That is, we can discover if it makes a mistake for the current agent. Did it mistake it for the agent in the last action suggesting that the memory persevered or did it mistake it for an agent that was not stored in working memory, thus suggesting more of a confusion type paradigm. This then will help us deduce if the model makes a mistake based upon a memory pattern stuck in working memory, or from some other source of error.

2.1.6 GABAergic memory field control

As has been discussed, following the execution of an action, working memory activity in dlPFC is reduced. We have hypothesized that inhibitory interneurons particularly of the fast spiking type would be idea for this task. Other models of working memory have also used a suppressive device to halt or wipe working memory (**dudes**). In our model, working memory is destroyed when an expectation is fulfilled. This happens though, just after this memory can be saved. This occurs for several reasons. The first is that following the successful completion of an action execution, if you want to complete another action, it makes sense to free up the memory used. Second, you need a brief delay in the memory destruction, just long enough for a gate to open for the memory to be stored. One avenue not explored is the violation of expectation. This may also signal that a working memory should be stored and destroyed following a signal

from anterior cingulate cortex which is believed to play a role in the detection of expectation violations (**dudes**). Thus, what we are exploring is one paradigm of a manner in which a working memory may be analyzed or cleared, in this instance if an expectation is fulfilled.

2.2 Model Details

One can view a layout of the model in figure 2. For every equation, output firing rates are approximated with the standard sigmoid (eq 1).

$$(1) \qquad S(x) = \frac{1}{1 + e^{(-\beta(x - \textit{offset}))}}$$

To start, we should note that each set of equations are compartmentalized in modules for the Neural Simulation Language, which we will describe later. At the top, we have created two recurrent memory fields, one for agent and one for action. From studies of memory fields we have hypothesized that these neurons are serotonergic 5-HT2a pyramidal neurons (Williams, Rao, Goldman-Rakic, 2002; Grace, 2000). The intended action/agent are placed in simulated leaky integrator neurons (eq 2 and 3). τ is the time constant with vm_i as the potential and Vm_i as the output from this neuron after the firing rate is estimated from the sigmoid. um_i is the input stimulus and pm is the inhibitory input from a GABAergic interneuron. In this case a chandelier neuron. All weights are represented with a wm.

Working Memory Neurons

$$(2) \qquad \tau m \frac{dvm_i}{dt} = -vm_i + Vm_i \bullet wm_0 + um_i \bullet wm_1 - pm \bullet wm_2$$

$$(3) \qquad Vm_i = S(vm_i)$$

Here the memory stays until the memory has completed leaking or is squelched by a strong suppression from fast spiking chandelier GABAergic interneurons. Action can be held in working memory until a tonically inhibitive interneuron gate slows, thus allowing the action to be released (eq 4-6). The basic idea behind the gate is that suppression is based upon tonic firing from term wt_1 this is reduced by an input from pt. The output is used to squelch input ut_i based upon the neurons output Vt. That is ut_i is gated by raising or lowering the tonic firing from Vt.

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

Tonic Gating Neurons

(4) $$\tau t \frac{dvt_i}{dt} = -vt + pt \bullet wt_0 + wt_1$$

(5) $$Vt = S(vt)$$

(6) $$Ut_i = ut_i - (ut_i \bullet Vt)$$

A feedback is received after a delay for the action executed. The action observed is compared with the action expected in a dopamine correlator module. The correlator acts as a fuzzy "and" circuit in that the more similar two incoming firing rates are, the more they are let through and not suppressed. The correlator is broken into two units. The first unit described in (eq 7-9) is a reciprocal inhibition circuit. In this formula ur_i^1 is this paths input. Vr_i^2 is the output from the reciprocal path. Again wr are weights for each term. In essence (7) is a GABAergic neuron with an input connection ur_i^1 and a reciprocal inhibition from the other GABA neuron Vr_i^2. The output Vr_i^1 is used to inhibit the axonal throughput of ur_i^1 as is described in (9). This happens by subtracting (increasing) inhibitory terms from this paths GABAergic neuron Vr_i^1 and the other paths GABAergic neuron Vr_i^2. The output Ur_i^1 is thus merely ur_i^1 suppressed by Vr_i^1 and Vr_i^2. A drawing of this circuit can be seen in figure 2.

Reciprocal GABAergic connection

(7) $$\tau r_i^1 \frac{dvr_i^1}{dt} = -vr_i^1 + ur_i^1 \bullet wr_0 - Vr_i^2 \bullet wr_1$$

(8) $$Vr_i^1 = S(vr_i^1)$$

(9) $$Ur_i^1 = ur_i^1 \bullet wr_2 - Vr_i^1 \bullet wr_3 - Vr_i^2 \bullet wr_4$$

Further the output is collected in second unit (eq. 10 – 12). This is basically an integrator that collects the output from the two paths ur_i^1 and ur_i^2 into ua_i^1 and ua_i^2. This is then summed at (12) to simplify the output that is sent to GABAergic neurons.

DA1 summation

$$(10) \qquad \tau a_i \frac{dva_i}{dt} = -va_i + ua_i^1 \bullet wa_0 + ua_i^2 \bullet wa_1$$

$$(11) \qquad Va_i = S(va_i)$$

$$(12) \qquad D = \sum_i Va_i$$

For the inputs the order of the feedback and feedforward action into the correlator unit are assumed to be in the same order to simplify the model. If an action working memory neuron fires at roughly the same rate as a neuron firing for the action observed then their firing is said to correlate. If enough correlation occurs, then a putative signal is sent by dopaminergic neurons to chandelier neurons at the memory fields destroying the memory. These can be seen in (eq 13, 14). u is the input from a correlator. The output is divided in (14) among several neurons in a memory field. This suppression is then seen in *pm* in equation 2.
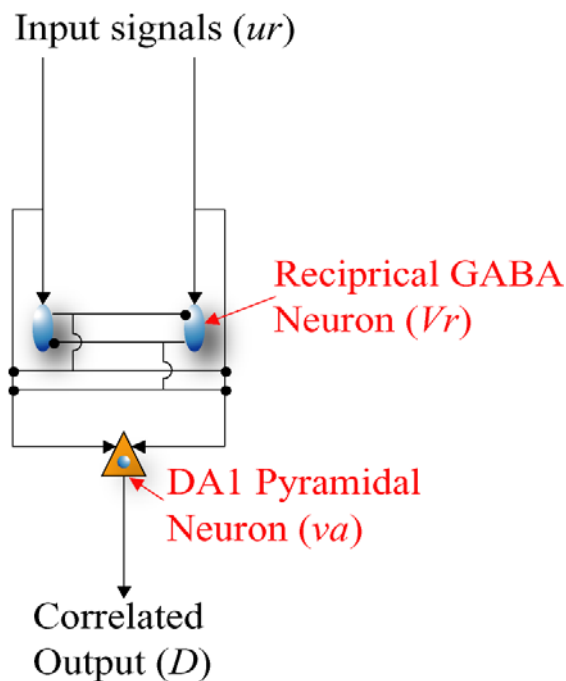

Figure 2. The correlator is basically a fuzzy "and" gate for associating two inputs. The more two inputs match in firing rate the more higher the firing rate from the DA1 neuron at the bottom. This is the basic unit for creating associations in the network.

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

Chandelier Neuron

(13)
$$\tau c \frac{dvc}{dt} = -vc + uc \bullet wc_0$$

(14)
$$Vc_i = S(vc)/Nc$$

However, just before that can happen, tonically inhibitive gates open allowing memory storage of the action observed with the expected agent from working memory.

2.3 Rationale

The three neuron types used are the three types that are hypothesized to play a role in schizophrenia (Volk, Lewis 2002; Beasley, Zhang, Patten, Reynolds, 2002; Koh, Bergson, Lidow, *et al*, 2003; Dean, 2003; Eastwood, Burnet, Harrison, *et al*, 2001; Aghajanian, Marek, 2000; Grace 2000). Memory fields are comprised of 5-HT2a type neurons since they have been shown to play a major role in working memory. For instance, research has shown that increasing the affect of 5-HT in dlPFC working memory increases its ability to hold memories and additionally increases the clarity of the memories (Williams, Rao, Goldman-Rakic, 2002).

Dopaminergic neurons are also hypothesized to play a role in dlPFC working memory (Seamans, Gorelova, Yang, et al, 2001; Durstewitz, Seamans, Sejnowski, 2000(1); Durstewitz, Seamans, Sejnowski, 2000(2)). The connections within the correlator are similar to (Durstewitz D, Seamans J K, Sejnowski T J, 2000(1)). Here, the major difference is that reciprocal connections from GABAergic neurons do not connect into the source neurons soma to control its firing rate, but instead they connect axonally to control throughput. This gives the circuit it fuzzy "and" ability. Also, direct reciprocal connections between source neurons are absent. The correlators also form modulatory connections to other pyramidal neurons via GABAergic interneurons (Tanaka, 2000; Durstewitz, Seamans, Sejnowski, 2000(1)). This creates a reciprocal connection between dopamine, GABA and serotonin neurons that has been observed in PFC. This is also a logical place for a dopamine neuron since it is strongly implicated in temporal difference correlation in other brain regions (Suri, Bargus, Arbib, 2000) as well as in reinforcement and associative learning (Waelti, Dickinson, Schultz, 2001; Suri, Bargus, Arbib, 2000). Thus, the dopamine neurons are able to fulfill their

role of creating associations that correlate an incoming action with the one held in working memory even if timing is not exact for the input. However, our model does not take advantage of the temporal difference properties at this time. Additionally, the dopamine neurons connect to and control the gating mechanism for memory storage. Such a mechanism has been proposed (Grace 2000). However, in our study we use a tonic GABA firing mechanism which could be found at (**where**) (**dudes**)[8], but the tonic mechanisms we employ are compatible with their study.

Chandelier neurons in dlPFC are used do to their high firing rate as well as the fact that they require very few putative inputs (Krimer, Goldman-Rakic, 2001). Chandelier neurons are also of interest since several studies have shown that they are deficient in dlPFC in patients with schizophrenia (Volk, Lewis, 2002; Beasley, Zhang, Reynolds, *et al* 2002). They have also been shown to play a major role in regulation of working memory in dlPFC (Krimer, Goldman-Rakic, 2001**;** Seamans, Gorelova, Yang, et al, 2001; Durstewitz, Seamans, Sejnowski, 2000a; Durstewitz, Seamans, Sejnowski, 2000b) and are also a fast spiking interneuron which is the major type of interneuron modulated by DA (Gorelova and Yang 1998) and provide most of the inhibition to the neocortex including the PFC (Kawaguchi and Kubota 1997). Thus, the model uses neurons that are known to play a role in working memory and use known patterns of connection if they have been found to exist.

The model of neurons also is important because in our experiment we need to attempt to account for the interaction of the three major neural transmitter types. Since the exact connections of these neurons is still yet to be determined, the model uses known connections where available and a best estimate where information in vague. Thus, while DA1, 5-HT2a and parvalbumin type neurons are known to interconnect and play a role in schizophrenia, the exact pattern of interconnections is not known. We have thus hypothesized that dopamergic neurons would be the best for an action correlation mechanism that provides feedback to the network since it has been shown to be important in associative learning activities. However, the precise connections created in the correlator, while feasible, are our best estimate for how such a circuit might work. Additionally, the direct feedback to GABAergic interneurons that control the working memory

---

[8] I'm not sure this exists in this brain region or not. This is for sure a weakness.

in our model is the simplest way we could surmise for a working memory control. This was necessary since as it has been mentioned, it has been shown that completion of an action removes observed working memory fields. GABAergic parvalbumin type interneurons in our circuit are used mainly due to the observation that they have a very strong inhibitory ability that could wipe a neural network very quickly (Volk, Lewis, 2002; Krimer, Goldman-Rakic, 2001).

2.4 Modeling Method

The model was created and implemented using the Neural Simulation Language *NSL* (Weitzenfeld A, Arbib M A, Alexander A, 2002). NSL is a modularized intermediate level neural simulator. It allows for large networks of neurons to be connected and simulated, but does not demand high precision details such as potassium and calcium ion gating properties. This allows the simulation to encompass a larger network thus allowing one to investigate larger neural networks. However, to do this, the workings of neurons are simulated using well-established mathematical approximation methods such as sigmoids for firing rates (Koch, 1999).

For the model, modules are created that represent chandelier neurons, 5-HT2a working memory neurons, DA1 correlators, and fast spiking GABA gating neurons. These are treated as the basic working memory model. This is then connected to a longer-term memory module comprised of correlators that collect the output from the working memory model and compare it to simulated longer-term memory. A decision about the identity of the incoming memory is based upon comparison between longer-term memories and the incoming working memory. The strongest correlation creates the best match; this is decided using the NSL *MaxSelector* model (Amari, Arbib, 1977). The long-term memory that correlates best with the incoming memory is selected as the correct memory using this method. This allows us to observe which agent is selected with what action. Here we can then determine which action and agent are being bound together and whether they are the correct one or are in error (figure 4).

3 Experiment

The primary goal of the experiment was to observe the working of the neural model we created in both healthy and diseased states. Thus, the baseline model was compared with models that elevated or decreased expression of 5-HT2a, DA1 and GABA in target neurons. This was done by increasing or decreasing the efferent or afferent gain to the neuron. Further, after diseasing neurons, treatment was applied by simulating the application of serotonin and dopamine antagonists that should mimic the actions of drugs used to treat schizophrenia. Thus, the idea was to create a schizophrenic simulation and observe what simulated actions would bind with what agents. Then we observed how a treated brain would react and possibly recover from its abnormal state into a less psychotic state.

For the diseased state, neurons were altered according to observations about an abnormal morphology if available. If none was available, a general gain was applied to the neurons in question at a level where change was noted if any. For instance, observation of abnormal morphology in Chandelier parvalbumin type GABAergic interneurons suggests that problems arise at the afferent connections to pyramidal neurons (Lewis, Volk, 2000). To address this, a gain was applied at the afferent stage in the GABAergic module. However, since research about dopaminergic neurons is more uncertain, they were tested by simply applying a gain to the neural potential thus, in general, increasing or decreasing spiking rates.

To create the simulation of a residual memory, each experiment used two simulated actions performed by the model. That is, the simulation will create a working memory for an action with an agent bound to it. The action will be simulated in time and completed, then a second action will be executed with another agent. Since the state of the model is reset between each experiment, the order in which experiments are carried out did not matter.

In essence the experiment was analogous to having a person execute or imagine two actions in sequence. For instance, the idea was for the computer to create the intention of an action with the intended agent, then execute that action, observe the feedback and make a decision about what action it observed and what the agent was that it expected to be bound with that action. Then it would a wait a few second, and then create

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

a new action/agent intention pair. Thus it would be like a person saying "how's the weather" and then thinking "the weather is rotten" with a corresponding intentional sequence *I say to someone* and then *I think to myself*. The idea is to catch the system matching up something incorrect like *I think to someone*.

Five possible memories for actions and five separate possible memories for agents were created to compare the model output to. As an example, the first memory is of action/agent *A* and the second memory is of action/agent *B*. The other three control sets of longer-term memories are created semi-randomly. Each memory and accordingly the code that represents each action and each agent are made up of five numbers
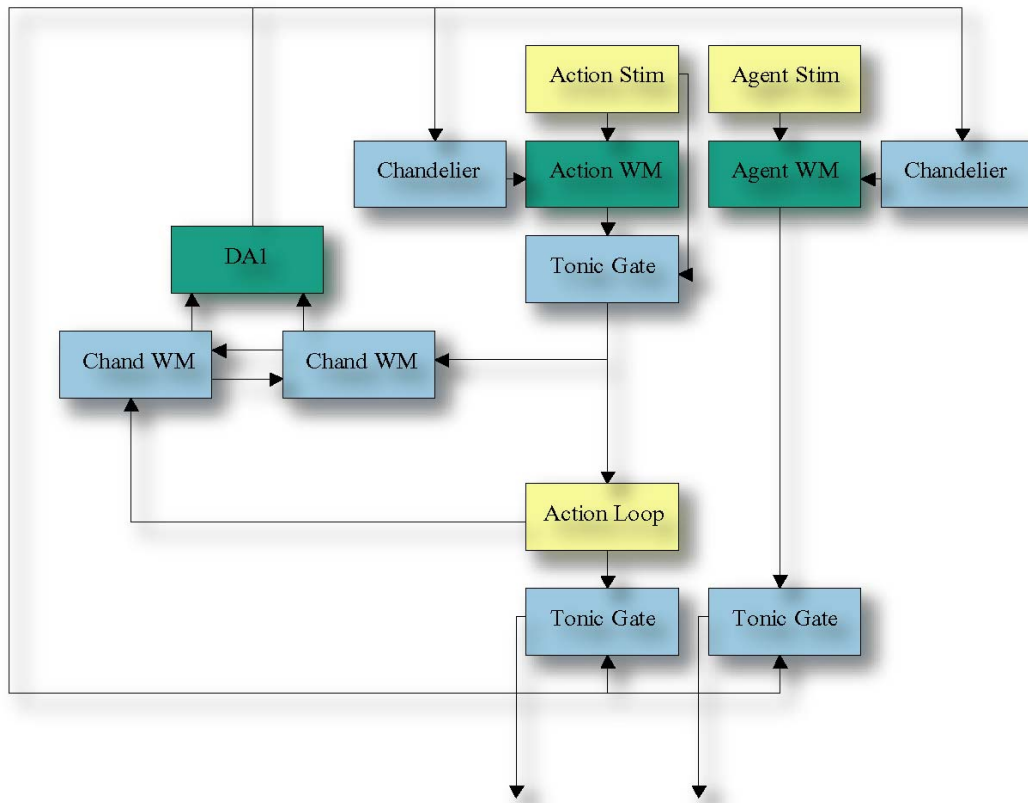

Figure 3. The NSL layout for the model seen in figure 1, creates several modules that connect to each other. This is a graphic of the working memory module. Yellow blocks indicate stimulus blocks that input some sort of stimulus into the network. Blue inhibitory interneurons act as tonic gates, correlators (chand WM) or as working memory wipers (Chandelier). Green blocks represent pyramidal neuron modules. Input from Action/Agent Stim is stored in Action/Agent WM and can be released into longer-term memory at the bottom by tonic gates. The DA1/Chand WM blocks act as a correlator between Action WM and input from the Action Loop. Chandelier neurons suppress WM when activated by the DA1 block.

ranging from 0.1 to 1.0. That is, each working memory field as well as each longer-term memory field is made up of five neurons that fire at a different rate depending on the input number, for instance, 1.0, 0.5, 0.2, 0.5, 1.0 could be a code for one working memory. This means for instance if the agent is me, I might be represented by a code like 1.0 0.8 0.5 0.8 1.0. If an incoming working memory matches 1.0 0.8 0.5 0.8 1.0 then, I believe that the intended agent was me. If not, and it matches to another memory, then I believe that that was the intended agent. Additionally, it is important to reiterate our theory that this particular system is egocentric, so we have two agents one of which must always be me and the other which can be any agent.

The process for when a working memory is created for an action, is that, it is held for a simulated interval of about one and a half seconds before it is executed when a tonically firing gate is suppressed. An interval of about one to two seconds is used to simulate time for feedback. This takes into account time for action planning, execution and observation. When the action is received back it is observed for two seconds. Following that, action observation is terminated. After that another interval of one second passes before another agent and action are placed into working memory. The second action follows the exact same timing rules for execution and observation.

The resolution of the model is set to 10 ms. this allows for neurons to spike as fast as 100 Hz if needed. For approximation of neuron spiking, a sigmoid is used. Euler integration is then used by NSL to create the activity of leaky integrator neurons in discrete computation. The MaxSelector model uses firing rates simulated by step and ramp functions. The simulated firing rates of all neurons in the model fall between 20 and 100 Hz.

Disease conditions are created by simulating the alteration of the expressions of several neurotransmitters. 5-HT2a is altered by increasing the weights of EPSP's to efferents in working memory neurons as well as recurrent connection weights by 5%. This number was chosen since it was the minimal value at which change was noted in the network. Since data on the morphology of 5-HT2a is uncertain, an exact number was intangible. GABA for chandelier neurons is altered by reducing its IPSP afferent weights to working

memory serotonin neural units by 25% in line with observed levels of decreased GAT-1 expression in these neurons (Volk, Lewis, 2002). IPSP's were not reduced for gating interneurons in this experiment. Even though they are fast spiking interneurons, we wanted to limit the scope of areas investigated and wanted to consider that another type of neuron may be acting as a gating mechanism (Grace 2000). Efferent EPSP's were increased into DA1 pyramidal neurons by 25 or 40%. Like with 5-HT2a neuron units, the exact morphological abnormality is still unclear so the adjustments applied reflected a low boundary to where changes were noted. It should also be mentioned that each of these three conditions were executed separately with the exception of treatment conditions. That is, for instance, dopamine neurons were altered without any alteration in 5-HT2a or GABA. In the treatment condition however, additional neurotransmitters were altered to simulate changes that are hypothesized to take place.

After alteration of the neural expression, treatment conditions were simulated by altering the expression of DA1 or 5-HT2a in the presence of the prior alterations. GABAergic expression treatment was not examined since major drugs used to treat psychosis are antagonists are dopamine and serotonin but not GABA. The treatment alterations were decreases in 5-HT2a by 5% and DA1 by 40%, again these numbers were chosen since it was at these levels that changes were noted in the network.

Experimental alterations were not performed on neurons in the longer-term memory module by alteration of neural expression in the same way as was done on working memory neurons. This was because they are not the focus of this experiment. However, this is not to say we do not feel they play a role in schizophrenia, it is instead due to the need to limit the scope of this work.

From the experiment the agent and action memory matching was determined by observing which memory gave the largest peak from a correlator. That is, the longer term memory that matched up better with the working memory input gave the highest neural firing rate, and thus, it gave the highest peak on a firing rate chart (figure 4).
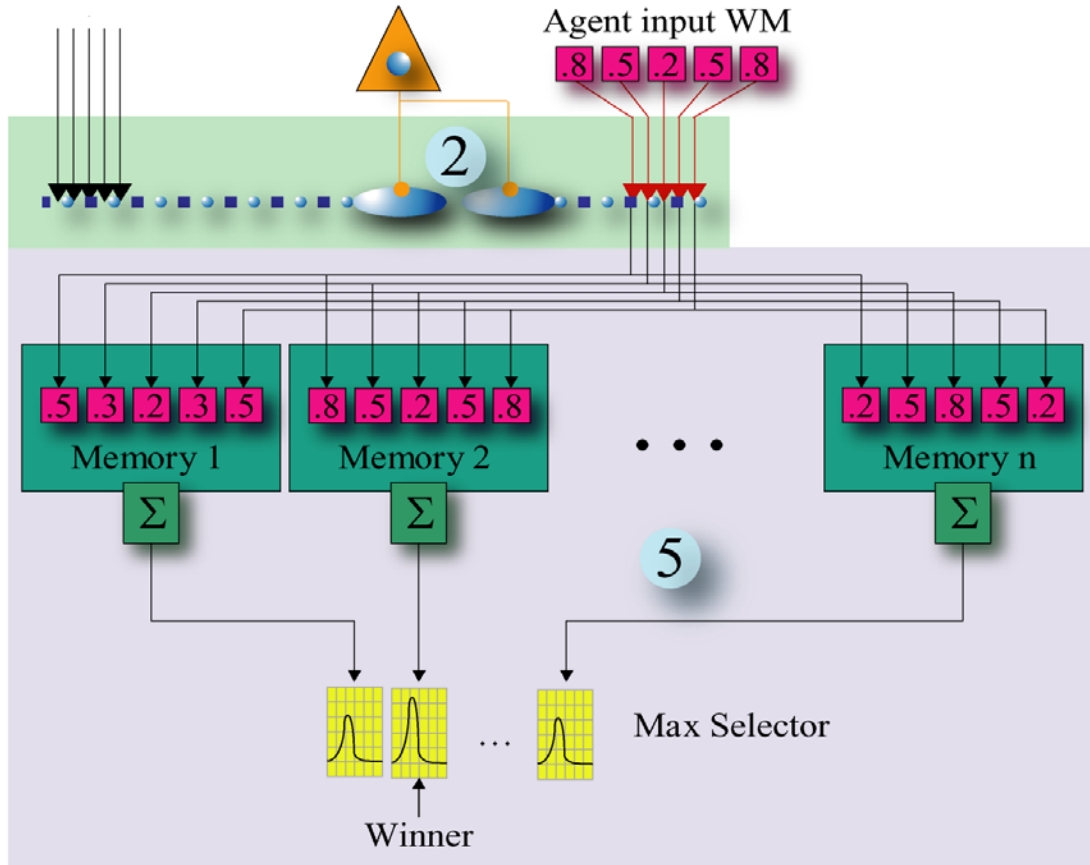
Figure 4. This is the longer term memory from figure 1 in detail. Coded inputs are feed in at the top (for instance .8, .5, .2, .5, .8). This memory is then matched against several other memory patterns already held in longer term memory using the standard correlator circuits described in figure 2. The memory that matches the strongest has more outputs from the correlator circuits. This creates a higher output firing rate. The memory selected as the correct one is the one with the highest firing rate as determined by the *MaxSelector* model.

The experimental results were classified by one of three criteria. The first is that the correct working memory matched up correctly with its corresponding longer-term memory representation. This means that if action *A* was committed then action *A* was observed correctly. The second outcome memory perseveration, was if the second action or agent was mistaken for the first. For instance, this is where action *A* is committed then action *B*, but the perception is that action *A* was committed twice. The third type of outcome, memory confusion is where a working memory is completely confused for some other object in longer-term memory. For instance, actions *A* and then *B* are observed, but one or more is confused for another action *C* which is held in longer term memory but was not executed. Thus from this, each experiment can have several outcomes. The first and second action/agent can be correctly identified by the

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.
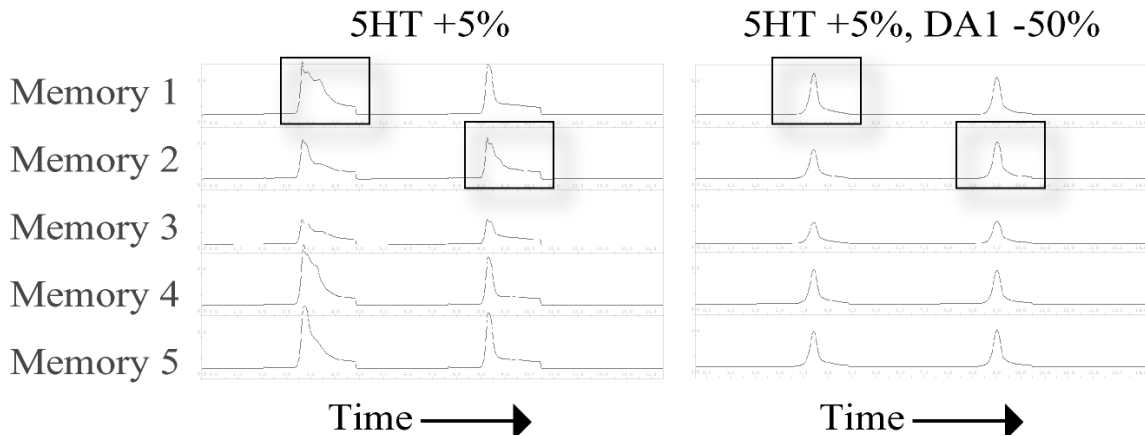

Figure 5. Outputs are shown from the simulation of two conditions, one a high 5-HT condition and the other a higher 5-HT condition treated with a dopamine antagonist. Two actions are executed and the spikes highlighted with the box is the output signal that should correlate the most, that is memory 1 with agent 1 and memory 2 with agent 2. As can be seen the untreated condition shows the ability to both persevere and confuse in working memory tasks. The treated brain, right, does not show memory confusion.

system, a second action/agent can be mistaken for the first and additionally an action/agent can be mistaken for one that did not happen in the experiment but was accessible as it was stored in longer-term memory.

4 Results

4.1 5-HT2a alterations

With an increase in the expression of serotonin working memory neurons by 5% both memory perseveration and memory confusion were observed. Thus, for agent working memory the second agent was mistaken for the first. Also, both the first and second agent could be mistaken for another unrelated

Experiment Condition	Memory Perseveration.	Memory Confusion
Base Line	No	No
5-HT(2A) +5%	Yes	Yes
5-HT(2A) –5%	No	No
5-HT(2A) +5% DA1 –50%	Yes	No
GABA +25%	No	No
GABA –25%	Yes	Yes
GABA –25% DA1 –25%	No	No
GABA –25% 5-HT(2A) –5%	No	No
DA1 +25%	No	No
DA1 –25%	No	No
DA1 +40%	No	Yes
DA1 +40% 5-HT(2A) –5%	No	No

Table 1 Each experimental condition is listed at the right. Also listed is any observed memory abnormality such as memory preseveration or memory confusion. Treatment conditions are highlighted in gray.
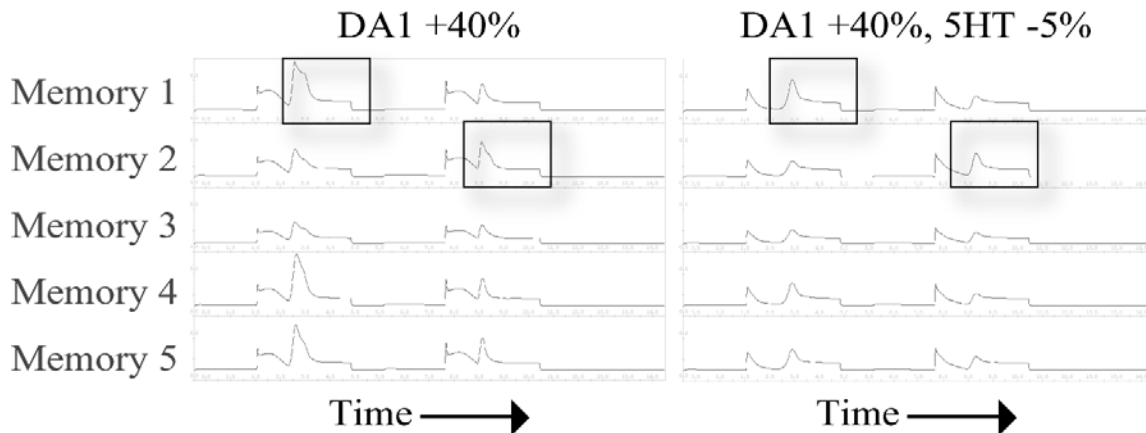
Figure 6. Dopamine abnormal condition on the left shows memory confusion but not perseveration. Treatment with a serotonin antagonists on the right seems to rectify the error.

longer-term memory agent. Reduction in 5-HT2a expression did not increase error, but a decrease by as much as 10% caused the firing rate of the outcome correlation to be virtually flat. Thus, no memory could be matched at this low level. Treatment by reduction of dopamine expression in DA1 correlation neurons by 50% alleviated some of the symptoms. Some memory perseveration was observed, but memory confusion was not. Thus, the model was induced into schizophrenia by the increase of 5-HT2a expression, but was partially treated by a simulated dopamine antagonist (table 1, figure 5). The result being that agents would not persevere in working memory, but a working memory for an agent could still be confused for another agent held in working memory.

The reason for these results is due to stronger holding by working memory units that create residual memories that incur on the second memory. Lowering DA1 expression treated the problem to a certain degree by letting less signal past the tonic gate into longer term memory. That is, correlation between action stored and action recognized had to be higher in order for a signal to get through the gate. This made the system more selective in the face of more ambiguous data. However, this decreased the ability for DA1 neurons to modulate the chandelier neurons that erase the working memory, but not by enough to overcome the gain of having a more selective tonic gate.

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.
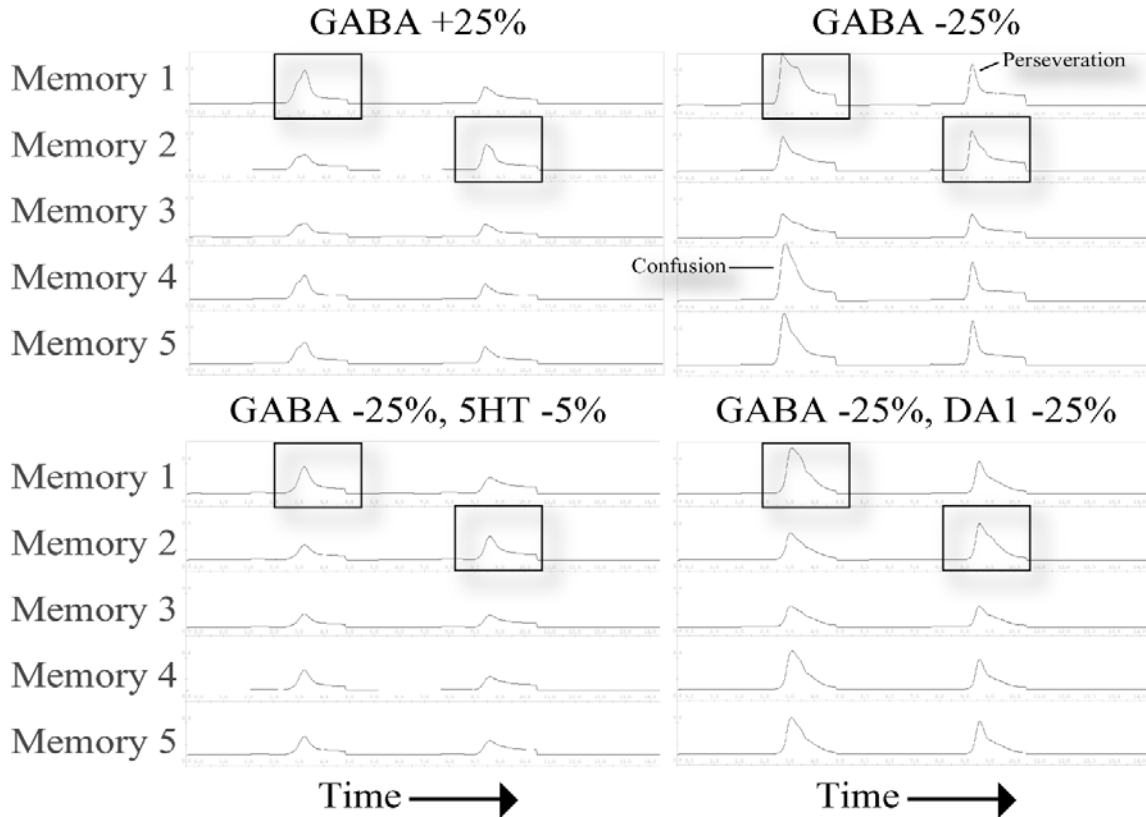

Figure 7 The NSL correlation for two action/agent executions compared against five different memories . The higher the peak the greater the correlation. Increasing GABA expression by 25% yields no difference in attribution for agent. Reduction by 25% shows both perseveration and confusion. Treatment with both dopamine and serotonin antagonists treats this problem.

4.2 DA1 Alterations

An increase or decrease in dopamine DA1 expression in the working memory correlators by 25% did not create any agent attribution error. An increase by 40% did not create perseveration errors but did increase the likelihood of making a confusion error. This error was counter balanced by a decrease in 5-HT2a expression by 5%. The results had some slight ambiguity in that firing rates had greater entropy, but perseveration and confusion errors were not present in the 5-HT2a treated condition (figure 6).

Increase in DA1 seemed to create attribution errors due to an increase in the ability for memories to passed through the tonic gates that had lower correlational values. That is, action observed was more likely to correlate with the action working memory, which ramped the signal up that passed through the tonic gate, making it more likely to resemble other memories.

T. Nathan Mundhenk, Michael A. Arbib

4.3 Chandelier Alterations

Reduction in GABA expression by 25% caused both memory perseveration as well as memory confusion. An increase by 25% did not cause any errors. Treatment by reduction in DA1 expression by 25% eliminated both perseveration and confusion errors. Reduction in 5-HT2a by 5% also had this effect. The reason for these results is the same as that for the 5-HT2a expression alteration. Working memory fields become easier to wipe with less 5-HT2a expression. Further, reduction of DA1 expression increased the level at which an action observed and an action executed needed to correlate. (figure 7)

5 Discussion

5.1 Auditory Verbal Hallucination Pathway

The model was able to reproduce errors in judgment about an action or agent when known or hypothesized neural abnormalities were created in the model. This demonstrates the feasibility for the idea that a residual memory caused by abnormal retention of working memory can lead to misattribution of an action to an agent in schizophrenia. Further, if a verbal thought action can be confounded with a verbal utterance action, it demonstrates how schizophrenic patients can experience sublingual vocalization as the decision for the course of action, of thinking verses speaking, is confused at a half way point due to lay over residual memory. This then means that the action for speaking could be caught between itself and an action for thinking when the thinking action creates a residual and leaves what should be a pure speaking action half way between speaking and thinking.

A feature not modeled here that would further enable auditory verbal hallucinations is that verbal thoughts and verbal utterances are produced and feedback in the same pathway. This idea similar to one modeled by Frith (Frith, Blakemore, Wolpert 2000), which poses the problem as a feedforward/feedback error. However, for auditory verbal hallucinations to occur, a verbal thought is released to the same areas of the

brain as a verbal utterance. Then the verbal thought returns to the same area in a feedback loop as a verbal utterance (figure 8). Such an idea is supported by research by (Stephane, Barton, Boutros, 2001) that showed that verbal production centers are heavily activated during auditory verbal hallucinations. How our model would fit in with such a paradigm is that an initial feedforward verbal intention is created. It travels to language centers for instance Broca's area so that it can be fully realized. That is, a prototype about a language thought is created in or around PFC and is sent to Broca's area as well as other language areas so that it can be turned into a true sentence in ones mind. In schizophrenic patient, this can become an auditory verbal hallucination because when the sentence feeds back, the action working memory does not account for it being a thought. Instead the working memory confounds it as an utterance. Misattribution comes when working memory of the agent is also confounded. In this way, a verbal thought is confused as speech. Further, the agent of the speech is confused. Here a schizophrenic patient *knows* the agent, or at least they think they know. This distinction helps differentiate schizophrenia from other disorders of involuntary speech and action production such as Tourette's disorder and alien hand sign. This is important since in such diseases, involuntary actions are not attributed to aliens or other agents but are instead known to the agent as self produced. For a healthy person we hypothesize that general feedforward/feedback violations of expectation can be handled, even if one forgets that one executed an action. What makes schizophrenia unique is that exceptions are not handled properly since it cannot be considered that an unknown involuntary action was in fact unknown if the agent is believed to be known.

It is important at this point to clarify the distinction between our model and other similar models. For instance, our model shares many similarities to the model by Frith (Blakemore, Oakely, Frith, 2003; Frith, Blakemore, Wolpert, 2000). For instance both models depend on a breakdown in the feedforward/feedback loop for action and perception to create misattributions of agent and action. However, our model depends on holding of a false working memory through residual rather than the lack or absence of a memory or alternatively we also believe that it is possible that misattribution comes from a binding mismatch of another kind as a result of increased correlational activities in dopaminergic neurons. Our theory can be defended through examination of neuropsychological evidence. For instance, (Williams, Rao, Goldman-Rakic, 2002) showed that increasing serotonin in working memory fields increased the firing rate and

Figure 8 The idea of an action and agent are created and sent to parietal areas to be sequenced into language or grasping movements. This includes both overt and imagined actions. Feedback is sent for both overt and imagined actions through the thalamus into the basil ganglia. Part of the model may rest in the anterior cingulate cortex since it is implicated in expectation operations. If the action is perceived external from expectation then it is experienced as an external verbal utterance.

clarity of memories. Since schizophrenia is treated effectively by atypical antipsychotics that are known serotonin antagonists (Merck Manual,1992) it is easier to imagine that working memory fields are over activated and not underactivited in schizophrenia. Further, neural simulations by (Durstewitz, Seamans, Sejnowski, 2000a; Durstewitz, Seamans, Sejnowski, 2000b; Seamans, Gorelova, Durstewitz, Yang, 2001; Tanaka 2000) have shown that working memory fields increase in the ability to hold memories with higher dopamine DA1 or other pyramidal neuron levels and also increase in their ability to wall out noise as well as other inputs to themselves. Since the typical neuroleptics used to treat schizophrenia are dopamine antagonists this lends further credence to the idea that higher levels of dopamine are creating more holding in working memory and not less. The idea of memory holding can be further extended by that fact that chandelier GABAergic interneurons as well as other fast spiking parvalbumin expressing interneurons have been found to be deficient in post mortem studies in schizophrenia (Volk, Lewis, 2002; Beasley, Zhang,

Reynolds, *et al* 2002). This is pertinent since it is hypothesized that these neurons strongly control working memories from dopaminergic inputs (Krimer, Goldman-Rakic, 2001; Seamans, Gorelova, Yang, *et al*, 2001; Durstewitz, Seamans, Sejnowski, 2000a; Durstewitz, Seamans, Sejnowski, 2000b; Tanaka, 2000). If GABA is depleted then working memory may not be able to be erased properly. This again could create the residual memory hypothesized.

Residual memories may also help in explaining why auditory verbal hallucinations can have ~~gender (DUDES), as well as~~ a[9] unique identity that allows for conversation between more than one hallucinated agent about a patient (DSM IV, 1994). Also, if a person had schizophrenia and knew about it, it would make it easier to make an attribution of error if they had no memory of the will to execute. Instead, schizophrenic patients continue to experience vivid hallucinations even in the presence of the knowledge of their condition **(Chadwick, Sambrooke, Rasch, Davies, 2000).**

The theory of residual memories could be further elaborated on by examining the action of hallucinogenic substance such as LSD. A strong serotonin agonist, when administered to healthy persons, creates psychotic symptoms similar to schizophrenia (Merck Manual, 1992; DSM IV, 1994) And whose effects on neural systems are hypothesized to be very similar to what is observed in schizophrenia (Grace 2000). Further, hallucinations from LSD have residual effects. For instance, moving objects such as hands appear to have trailers or tracers to those who are under the influence of LSD **(DUDES)**.

5.2 Schizophrenia, pathways and the language ready brain

It has been asserted that language evolution involved the creation of the language ready brain (Arbib 2003; *in press*). The process for the creation of language involved the procession from being able to grasp objects with hands, as most primates can, into a signed language and then to spoken languages. While it not asserted that this procession is completely serial, it is asserted that the ability to grasp objects created the ability to sequence behaviors in a way that eventually lead to the ability to sequence words into language.

---

[9] Since I am having trouble finding a specific reference about gender we can strike this I guess.

The belief is that the area at the center of the evolution of the language ready brain is the human homologue to the monkey F5 which is believed to be Broca's area.
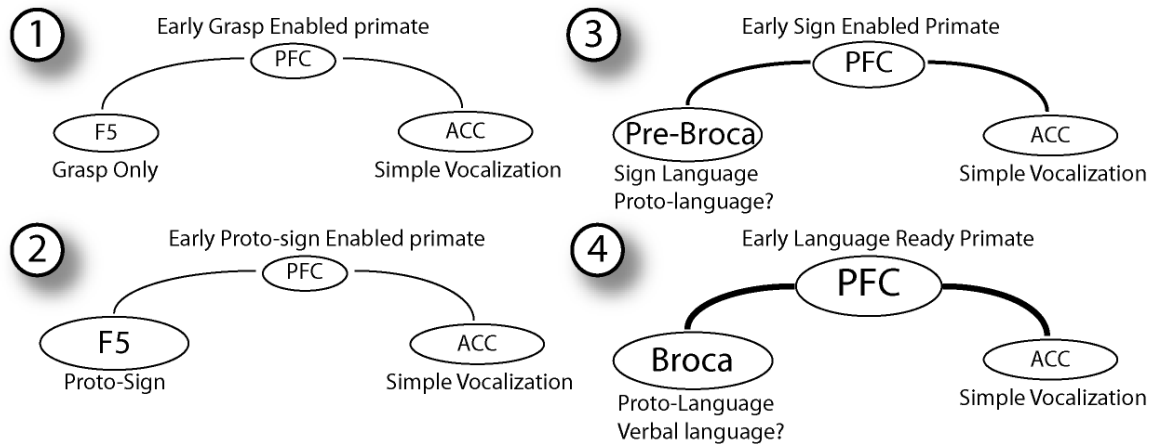

Figure 9 (1) Early pre-human primates had the ability to grasp and the will to create basic vocalizations. (2) Grasp evolved into proto-sign with greater ability placed on F5 and grasp related areas. (3) Growth in PFC linked the will to vocalize with the ability to create sign linking complex sign language centers with vocalization centers in ACC (4) The will to vocalize is completely connected to complex language center now Broca's area via PFC. Thus, vocal centers connected to Brocas syntax and parsing centers creating proto-language or the ability to have a full language ready brain.

In the monkey, verbal utterances are believed to be processed through the ACC (Jürgens, 1997) unlike in humans where processes have migrated to Broca's area. A key step in the evolution of the language ready brain thus is the observed movement of language centers from ACC to Broca's area. It is interesting to note that in humans both Broca's area and ACC may be connected via the PFC, which is highly evolved in humans when compared to non-human primates. Thus, the PFC may connect the will to communicate or the intention to create verbal utterances in ACC to the ability to sequence complex behavior in Broca's area. The PFC would then have become a conduit to help aid in the creation of the language ready brain by helping to connect these two areas.

In the evolutionary scenario we have presented, language that was spoken was created not only by a gain in F5's abilities to process grasping sequences but also through its eventual connection to early vocalization centers in ACC. A key was not just this linking but also the evolution of working memory to store the intention of communication. As such, PFC needed to evolve at the same time in order to store the intention behind the complex sequences involved in proto-sign. In order for a verbal language ready brain to be

created, the will to communicate verbally was linked from the ACC to PFC then to F5. Thus the verbal language centers in ACC, which had only very primitive abilities, gained access to what would become Broca's area.

The addition of PFC was necessary to language evolution in that the mirroring abilities of F5 are necessary, but not sufficient for language creation. The will and intent of language creations must also be handled and stored in order to create symmetry between agents. Thus, ACC may create this will and intent, which is stored in PFC. When an utterance is received back, from the world, the result can be compared against the intent from ACC using the working memory stored in PFC. This may help explain the observation that ACC is involved in handling exceptions to certain actions since it has an understanding of the basic intentions of the action. It could do this if it's an expert on the pattern of the intentions and is able to compare the incoming result with the intention stored in PFC.

Additionally, ACC may have provided a connection to other agents. Mirroring behavior that comes from F5 does not necessarily create a communicative connection to other agents. Thus, I may understand what another agent is doing, but this is not complete a basis for communication. In general I must identify another agent and send a message to that agent in a manner that I believe will be received and understood. It is also helpful if I know that they received the message. Thus, early verbal primate communication involved creating that link. Connection to ACC from F5 may have allowed an understanding of how to link to another agent in order to establish communication.

This then lends itself to the theory of schizophrenia presented thus far. The dlPFC becomes involved in a conduit between ACC and Broca's area that link the intention of what one would like to say and the method of communicative interaction from ACC to the ability to sequence and perhaps desequence the message. Thus, the dlPFC holds the intention of language execution along with agent information during the act of communication. It is thus an intermediary between the initial amorphic idea of communication and the sequenced final product.

The egocentric component talked of may also be a result of this marriage. This may be because non-human primates make only simple verbal calls from themselves to others. Thus, one agent is always the self. Functionally this may be in tact in the dlPFC. That is, ACC may originally have only dealt in interactions between the self and another agent. Thus, the bridge that was created had an egocentric construction. While non egocentric language and recognition centers have obviously developed in humans, their role has perhaps more to do with scene understanding than establishing a link for direct agent to agent communication. Therefore, it is not unreasonable to assert that these would be separate operations processed in different section of the human brain. That is, the reason behind the working memory we have been talking about it in part creating the mental link between the self and other agents. As such it does not need to process the linking between two other agents. Additionally, from the egocentric symptoms observed in schizophrenia, we believe it does not.

5.2 Action, locus of disorder and implications

As we have mentioned, we believe that the symptoms of schizophrenia may be related to either a temporal binding problem from correlation or it may be related to a temporal binding problem for a residual memory or abnormal memory holding. Our model was able to simulate both situations which creates several possibilities. The first is that both may be happening in patients with schizophrenia. This could be the case if another neurochemical not mentioned here was effected that underlies the action of more than just one of the neurons simulated. Thus, abnormal expression effecting dopaminergic neurons may lead to abnormal correlation of actions while abnormal expression in 5-HT2a may lead to a residual memory effect. Additionally, effects of dopamine or serotonin not modeled here may link serotonin to correlating activities or dopamine to basic storage activities. Thus, there are many possibilities. However, since dopamine is a better candidate for correlational activities it suggests that if temporal binding issues stem from correlation, abnormal dopamine expression should be found. We suggest that the best way to solve for this question will be to finally discover whether or not dopamine is indeed a culprit in schizophrenia. If not, we suggest that temporal binding for correlation is intangible. Additionally, the discovery that GABA or 5-HT is are more responsible would lend more credence to the hypothesis that an abnormal holding or residual memory effect is responsible.

5.3 Working memory and catatonic behaviors

The theory we propose also lends weight to the single spectrum hypothesis (**NAME?**) of schizophrenia (**DUDES**). The theory of a single spectrum states that what seems like sometimes-unrelated symptoms in schizophrenia for instance catatonia and auditory verbal hallucinations are all a part of one spectrum of the same disease. We support such an idea because if an initial working memory plan for action can become strongly held then you would expect to see stereo typed repetitive behaviors observed in the most ill of schizophrenic patients. That is, because the working memory for an action becomes almost completely stuck, a patient could them self become stuck in an action loop. The same action or essence of an action would become executed repetitiously. The single spectrum is then seen as the level of holding in working memory. Weak holding leads to auditory verbal hallucinations, agency misattribution and even confusion, while strong holding of working memories would lead to a catatonic state where new actions cannot be realized easily (**SITE SOME PAPERS**).

5.4 Working memory and agent attribution errors

It has been observed that schizophrenic patients have difficulty in determining the agent of an action. This error seems to increase with increases in observed error. For instance, in studies of self-attribution where visual feedback for hand movements of participants are altered, schizophrenic patients make more attribution errors as the image of their movements are altered both temporally and angularly (Franck N, Farrer C, Jeannerod M, *et al*, 2001). A higher level of erroneous self attributions could be made based upon both agency attribution errors as we have discussed and an insensitivity to errors which could be caused for instance by a faulty understanding of their own action. Either could be accounted for in our model. In the first possibility, errors could be made because the agent of action is held as a residual and misattribution is made based upon the residual agent memory. The second could happen if higher levels of dopamine increased the likelihood of making a correlation and an erroneous memory match.

Attribution errors in recalling if words were read aloud (Franck N, Rouby P, Geogieff N, *et al*, 2000) or thought to themselves can also be explained by residual action working memory. In the model we

presented, a residual of action could create the observation that a word was read aloud when in fact it was read silently and vise versa. This is because if a prior memory is of the action of reading aloud, it would pollute the working memory of reading silently. The new working memory polluted by the residual would be half way between both memories and possibly interpretable as either.

5.5 Implications for Hippocampal morphology

It has been suggested in many research papers that the hippocampus is the locus of schizophrenia. This is based upon studies that show a decrease in volume after disease onset. While our model focuses on the PFC as a more likely cause, it is possible that hippocampal disruption aids in the creation of schizophrenia. For instance, during memory classification and storage, it may check the input data for validity. If the hippocampus is disturbed, it may not be able to perform such a function. However, it is also possible that observed degradation of the hippocampus is secondary. That is, disorganization of the PFC creates data with more entropy that in turn causes atrophy to the hippocampus, which is connected, afferent to the PFC (Swanson 1981). It should also be noted that agency or the sense of agency may be more hippocampal in nature. For instance, since the hippocampus is believed to play a role in context processing (**dudes**), if the knowledge of agent is considered a context, then hippocampal role may be pivotal.

5.6 Dopamine, perseveration and treatment

In the model, decreasing dopamine expression was less effective at reducing confusion between misattribution of agent/action to the last action/agent due to residual. This is to be expected since dopamine antagonists used to treat schizophrenia are generally less effective than the newer atypical neuroleptics, which antagonize serotonin stronger. It could be expected then that any simulation that attempts to reduce dopamine levels as a treatment should see less effect when compared with serotonin based treatment.

5.7 Dopamine DA1 as suppressor and exciter in modulation

Studies have indicated a puzzling interaction between DA1 expressing neurons and GABAergic interneurons in that DA1 seems to modulate GABAergic interneurons to increase IPSP's and also decrease IPSP's in other GABAergic interneurons as dopamine is increased. We account for this in our model from

the need to simultaneously modulate gating neurons as well as memory wiping neurons. Thus, a DA1 expressing neuron may need to both activate some GABAergic interneurons and deactivate others at the same time in order to let a memory move to the next section in the brain while clearing it from this section.

5.8 GABA treatment prospects

While GABAergic interneurons may play an important role in schizophrenia it is puzzling that drugs meant to increase GABA expression seem to be far less therapeutic than drugs that decrease dopamine or serotonin. However, this may be because benzodiazapines that increase GABA expression may be far too broad. That is, abnormal expression in GABAergic neurons in schizophrenia have been found only in certain layers of the PFC and additionally only in parvalbumin expressing interneurons and to some degree calbindin interneurons, but not in calretenin interneurons. Thus, levels that would be needed to be effective in treatment of schizophrenia would overmedicate otherwise healthy interneurons. Thus, a balance of medication could not be found with these medications. This may suggest that drugs that target parvalbumin interneurons selectively would have a far greater possibility of success. Indeed, while treatment with benzodiazapines has not been demonstrated in schizophrenic patients, augmented treatment has been shown effective. In particular, the larger the volume loss in pre frontal cortex, the better patients tend to respond to treatment by alprazolam, a common fast acting benzodiazapine (Seeley W W, Turetsky N, Reus V I, Wolkowitz O M, 2002).

5.9 Locus of the disorder

Our experiments suggest that the dopamine, serotonin and GABA hypothesis are all plausible. The locus of the disorder could be in any one of these neural types. Discovering which one is responsible could conceivably be confounded by their reciprocal connections in that over or under activation from one neuron type could spill over to another. For instance, reduced suppression from GABAergic interneurons could create the observation that 5-HT expressing neurons are over active. Further, over activation of 5-HT neurons could conceivably result in future morphological abnormalities from plastic effects of neurons as they adapt to their over active state.

Since dopaminergic neural morphological differences are not well established in schizophrenia combined with the fact that dopamine antagonist neuronleptics are less effective than the newer atypical neuroleptics that are stronger serotonin antagonists suggest that the dopamine hypothesis may be less favorable. While drugs such as amphetamines can create psychotic symptoms at very high doses and are agonists of dopamine, these type of symptoms are usually the result of long term habitual abuse and not a single dose as one would expect in a pure dopamine paradigm (Merck Manual, 1992). Alternatively, psychosis from LSD can result from a single dose, which suggests a stronger role for 5-HT2a. Further, GABA can also be implicated since drugs that are agonists of it can cause sudden psychosis when the drug is withdrawn after long-term substance abuse.

It is also important to tease apart out models account of psychosis from models that rely on basal ganglia. Since the basal ganglia is an integral part of the language feedback pathway, it is conceivable that disruption of this part of the brain as well as many other brain areas could create psychosis. The argument for our model here comes from studies of agency in schizophrenia and not from other diseases where psychosis may be observed such as in advanced stage Alzheimer's and other diseases of mature onset dementia.

5.10 Relation to the MNS

Since our model has very similar treatment of auditory verbal hallucinations as well as delusions of influence, it lends credence to the theory that language and grasping are strongly bound. This is also bolstered by the fact that hallucinations and misattribution in schizophrenia are almost primarily related to language and hand movements as well as the feedback and observation of their actions (Check again). Since language and grasping are among the newest additions to the primate brain, it would not be surprising if they were more prone to a unique break down since evolution is still working the kinks out of the new system. Thus, many of the most profound diseases of the brain that seem to be narrowly system specific (Schizophrenia, Parkinson's disease, Huntington's Disease) effect complex movements such as grasping and language since the systems are new and require further evolution to weed out the propensity of such systems to break down so easily. **(Lots of sitations)**

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

Our model also helps explain the system for attentiveness observed in the Mirror Neuron System. For instance, when a Macaque monkey observes its own grasp F5 mirror system is activated. However, it is only activated when observing the grasp of another agent if the action of that agent is interesting, for instance if it picks up a raison (Rizzolatti, Fadiga, Gallese, Fogassi, 1996). Since it is our hypothesis that the working memory in our study is egocentric then it suggests a strong coupling with MNS. Additionally, since the working memory we have explored is responsible for a sense of agency and action in both grasping and language, we believe more evidence is provided for close interaction between language and grasp centers thus lending weight to the argument that language evolved from grasping.

5.11 The Agent is much bigger than this model

In our review it is important to note that while we feel the dlPFC plays an important roll in agency, we do not feel it is the only component. Indeed, agency may be mapable to many other brain centers. Here the PFC is involved in agency from its expectations, predictions or plans for an action. In the absence of prior executive knowledge of an agent and its action, other centers of the brain may have to piece together knowledge of agency. For instance, posterior parietal cortex **(???)** as well as the medial prefrontal cortex have been implicated in agency **(Decety Paper).** In such studies agency was tested by varying the predictability of an action as well as the actor of the action. From this, it can be hypothesized that more than one center in the brain may be responsible for agency attribution in general. However, in our model, agency attribution is committed when an agent and action are bound and expected. In the absence of such knowledge, our model makes no assumptions about the agent and action.

5.12 Model applicability to Attention Deficit Hyper Activity Disorder (ADHD)

The models performance also has suggestions for the course and treatment of other disorders. For instance it has been an enigma that amphetamines, which have strong stimulant properties, are effective at treating ADHD, which is characterized by hyperactivity and constant motor activity. It is conceivable that by increasing dopamine, as amphetamines are hypothesized to do, they increase the correlation between agent and action as in our model. This could create a greater self-awareness for actions than would not be

obtained otherwise if lower dopamine levels were preventing patients with ADHD from correlating their own actions to themselves.

6 Conclusion

Using realistic dorsal lateral prefrontal neural connections we were able to induce a psychotic state in our simulation through increasing DA1, 5-HT2a and parvalbumin GABA expression, which are all, hypothesized to play a role in schizophrenia. Since the model demonstrated psychosis from an abnormal residual memory effect we believe this, taken with other evidence of cognitive or biological functioning, suggests that a residual memory may be a possible cause for many of the positive symptoms observed in schizophrenia, particularly where agency attribution is concerned. We also conclude that supressionary effects, loss of memory, or memory inactivation for an action created do not explain positive symptoms in schizophrenia as well. Additionally, we suggest that the working memory found abnormal in schizophrenia is egocentric and is strongly linked to activities involving the mirroring of actions. This adds to the weight of arguments that language evolves from grasping since the egocentric working memory is strongly linked to both language and grasping agency attribution.

References

Aghajanian G K, Marek G J, 2000, Serotonin model of schizophrenia: emerging role of glutamate mechanisms, *Brain Research Reviews*, **21**(2000): 302-312

Arbib, M.A., 2003, From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics, Behavioral and Brain Sciences (*in-press*)

Arbib M A, Billard A, Iacoboni M, Oztop E, Synthetic brain imaging: grasping, mirror neurons and imitation, *Neural Networks*, **13**(2000):975-97

Beasley C L, Zhang Z J, Patten I, Reynolds G P, 2002, Selective deficits in prefrontal cortical GABAergic neurons in schizophrenia defined by the presence of calcium-binding proteins, *Biological Psychiatry*, **52**(2002):708-715

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

Bender S, Müller B, Oades R, Sartory G, 2001, Conditioned blocking and schizophrenia: a replication and study of the role of symptoms, age, onset-age of psychosis and illness-duration, *Schizophrenia Research*, **49**:157-70

Blakemore S, Oakley D A, Frith C D, Delusions of alien control in the normal brain, *Neuropsychologia*, **41**:1058-1067

Bowers (1974)

Bracci E, Centonze D, Bernardi G, Calabresi P, 2003, Voltage-dependent membrane potential oscillations of rat striatal fast-spiking interneurons, *J Physiol*, **549**(1): 121-130

Brébion G, Gorman J M, Amador X, Malaspina D, Sharif Z, 2002, Source monitoring in schizophrenia: characterization and associations with positive and negative symptomology, *Psychiatry Research*, **112**:27-39

Buttler R W, Jenkins M A, Sprock J, Braff D L, 1992, Wisconsin card sorting test deficit in chronic paranoid schizophrenia: evidence for a relatively discrete subgroup?, *Schizophrenia Research*, **7**:169-176

Carrilho P E M., Caramelli P, Nitrini R, *et al*, 2001, Involuntary hand levitation associated with parietal damage, Another alien hand syndrome, *Arq Neuropsiquiatr*, **59**(3-A):521-5

Chadwick P, Sambrooke S, Rasch S, Davies E, 2000, Challenging the omnipotence of voices: group cognitive behavior therapy for voices, *Behaviour Research and Therapy*, **38**(2000):993-1003

Dean B, 2003, The cortical serotonin 2A receptor and pathology of schizophrenia: a likely accomplice, *Journal of Neurochemistry*, **85**(2003):1-13

Diagnostic and Statistical Manual of Mental Disorders, Forth Edition, 1994, American Psychiatric Association, Washington DC

Dominey P, Arbib M, Joseph J P, 1994, A model of Corticostriatal Plasticity for Learning Oculomotor Associations and Sequences, *Journal of Cognitive Neuroscience,* **7**(3):311-36

Durstewitz D, Seamans J K, Sejnowski T J, 2000, Dopamine-Mediated stabilization of delay-period activity in a network model of prefrontal cortex, *Journal of Neurophysiology*, **83**(3):1733-50

Durstewitz D, Seamans J K, Sejnowski, 2000, Neurocomputational models of working memory, *nature neuroscience*, **3**(2000):1184-1191

Eastwood S L, Burnet P W J, Gittins R, Baker K, Harrison P J, 2001, Expression of serotonin 5HT-2a receptors in the human cerebellum and alterations in schizophrenia, *Synapse*, **42**(2001): 104-114

Fletcher P C, Shallice T, Dolan R J, 2000, "Sculpting the Response Space" – an account of left prefrontal activation at encoding, *NeuroImage*, **12**(2000):404-417

Franke P, Maier W, Hain C, Klingler T, 1992, Wisconsin card sorting test: an indicator of vulnerability of schizophrenia?, *Schizophrenia Research*, **6**:243-249

Franck N, Farrer C, Georgieff N, Marie-Cardine M, Dalery J, d'Amato T, Jeannerod M, 2001, Defective Recognition of One's Own Actions in Patients with Schizophrenia, *American Journal of Psychiatry*, **158**(3):454-9

Franck N, Georgieff N, *et al*, 2000, Confusion between silent overt reading in schizophrenia, *Schizophrenia Research*, **41**:357-64

Frith C D, Attention to action and awareness of others minds, 2002, *Consiousness and Cognition*, **11**(2002): 481-487

Frith C D, Blakemore S, Wolpert D M, 2000, Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action, *Brain Research Reviews*, **31**(2000):357-363

Garcia-Toro M, Blanco C, Salva J, 2001, Psychopathology and the binding problem, *Medical Hypothesis*, **57**(6): 718-723

Gao W, Goldman-Rakic P S, 2003, Selective modulation of excitatory and inhibitory microcircuits by dopamine, *PNAS*, **100**(5): 2836-2841

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

Grace A A, 2000, Gating of information flow within the limbic system and pathophysiology of schizophrenia, *Brain Research Reviews*, **31**(2000): 330-341

Haggard P, Martin F, Taylor-Clarke M, Jeannerod M, Franck N, 2003, Awareness of action in schizophrenia, *NeuroReport*, **14**(7): 1081-1085

Hazlett E A, Buchsbaum M S, Harvey P D, 2000, Hypofrontality in unmedicated schizophrenia patients studied with PET during performance of a serial verbal learning task, *Schizophrenia Research*, **43**(2000): 33-46

Inzelberg, R., P. Nisipeanu, S. C. Blumen, R. L. Carasso, 2000, Alien Hand Sign in Creutzfeldt-Jakob disease, *J Neurol Neurosurg Psychiatry*, **68**(1): 103-4

Jorgensen P, 1994, Course and outcome in delusional beliefs, *Psychopathology*, **27**(1-2):89-99

Jorgensen P, Jensen J, 1994, Delusional beliefs in first admitters. A clinical description. *Psychopathology*, **27**(1-2):100-112

Jürgens, U. (1997) Primate communication: Signaling, vocalization, in *Encyclopedia of Neuroscience*, Second Edition, Amsterdam: Elsevier.

Kestler L P, Walker E, Vega E M, 2001, Dopamine receptors in the brains of schizophrenia patients: a meta-analysis of the findings, *Behav Pharmaco*, **12**(5):355-71

Koch, C, 1999, Biophysics of Computation, Oxford University Press, New York

Kobayashi S, Lauwereyns J, Hikosaka O, 2002, Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex, *J Neurophysiol*, **87**:1488-1498

Koh P O, Bergson C, Undie A S, Goldman-Rakic P S, Lidow M S, 2003, Up-regulation of the D1 dopamine receptor-interacting protein Calcyon, in patients with schizophrenia, *Archives of General Psychiatry*, **60**(2003): 311-319

Krimer L S, Goldman-Rakic P S, Prefrontal Microcircuits: Membrane properties and excitatory input of local, medium and wide arbor interneuronrs, 2001, *The Journal of Neuroscience*, **22**(11): 3788-3796

Lanser M G, Berger H J C, Ellenbroek B A, Cools A R, Zitman F G, Perseveration in schizophrenia: failure to generate a plan and relationship with the psychomotor poverty syndrome, *Psychiatry Research*, **112**: 12-26

Leon M I, Shadlen M N, 1999, Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the Macaque, *Neuron*, **24**:415-425

Lewis R, Kapur S, Zipursky R B, *et al*, 1999, Serotonin 5-HT2 receptors in schizophrenia: A PET study using [18F]Setoperone in neuroleptic-naïve patients and normal subject, *American Journal of Psychiatry*, **156**(1):72-78

McDowell J E, Clementz B A, 2001, Behavioral and brain imaging studies of saccadic performance in schizophrenia, *Biological Psychology*, **57**:5-22

The Merck Manual, Sixteenth Edition, 1992, Merck Research Laboratories, Rahway, N.J.

Narr K L, van Erp T G M, Toga A W, *et al*, 2002, A Twin study of genetic contributions to hippocampal morphology in schizophrenia, *Neurobiology of Disease*, **11**(1): 83-95

Nieman 2000

Perlstein W M, Dixit N K, Carter C S, Noll D C, Cohen J D, 2003, Prefrontal cortex dysfunction mediates deficits in working memory and prepotent responding in schizophrenia, *Biological Psychiatry*, **53**(2003): 24-38

Petersson K M, Elfgren C, Ingvar M, 1999, Dynamic changes in the functional anatomy of the human brain during recall of abstract designs related to practice, *Neuropsychologia*, **37**(1999): 567-587

Posada A, Franck N, Georgieff N, Jeannerod M, 2001, Anticipating incoming events: an impaired cognitive process in schizophrenia, *Cognition*, **81**:209-225

Purdon S E, Woodward N D, Flor-Henry P, 2001, Asymmetrical hand force persistence and neuroleptic treatment in schizophrenia, *J Int Neurophychol Soc.*, **7**(5):606-614

Rizzolatti G, Fadiga L, Gallese V, Fogassi L, 1996, Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, **3**(2): 131-141

The agent I think I am: Modeling symptoms and observations in schizophrenia as they relate to a sense of agency and the ability to mirror behavior.

Rougier N P, O'Reilly R C, 2002, Learning representations in a gated prefrontal cortex model of dynamic task switching, *Cognitive Science*, **26**: 503-520

Schultz W, 2002, Getting formal with dopamine reward, *Neuron*, **36**:241-263

Seamans J K, Gorelova N, Durstewitz D, Yang C R, Bidirectional dopamine modulation of GABAergic inhibition in prefrontal cortical pyramidal neurons, 2001, *The Journal of Neuronscience*, **21**(10): 3628-3638

Serra A M, Jones SH, Toone B, Grey JA, 2001, Impaired associative learning in chronic schizophrenics and their first-degree relatives: A study of latent inhibition and the Kamin blocking effect, *Schizophrenia Research*, **48**(2-3):273-89

Stephane M, Barton S, Boutros NN, 2001, Auditory verbal hallucinations and dysfunctions of the neural substrates of speech, *Schizophrenia Research*, **50**(1-2):61-78

Suri R. E., Bargas J. Arbib M. A., 2000, Modeling functions of Striatal Dopamine modulation in learning and planning, *Neuroscience*, **103**(1): 65-85

Swanson L W, 1981, A direct projection ammons horn to prefrontal cortex in the rat, *Brain Research*, 217:150-154

Tanaka S, 2000, Roles of intracortical inhibition in the formation of spatially tuned delay-period activity of prefrontal cortical neurons: computational study, *Prog. Neuro-Psychopharmacol. & Biol. Psychiat*, **24**:483-504

Tanaka S, 2000, Computational approaches to the architecture and operations of the prefrontal cortical circuit for working memory, *Prog. Neuro-Psychopharmacol. & Biol. Psychiat*, **25**:259-281

Thompson P, Rapoport J, 2001, Mapping adolescent brain changes reveals dynamic wave of accelerated gray matter loss in very early-onset schizophrenia, *PNAS*, **98**(20):11650-55

Volk D W, Lewis D A, 2002, Impaired prefrontal inhibition in schizophrenia: relevance for cognitive dysfunction, *Physiology and Behavior*, **77** (2002): 501-505

Waelti P, Dickinson A, Schultz, 2001, Dopamine responses comply with basic assumptions of formal learning theory, *Nature*, Jul **412**:43-7

Weitzenfeld A, Arbib M A, Alexander A, The Neural simulation Language: A system for brain
modeling, 2002, The MIT Press, Cambridge, Massachusetts

Williams G V, Rao S G, Golman-Rakic P S, 2002, The physiological role of 5HT-2A receptors in
working memory, *The Jounral of Neuroscience*, **22**(7): 2843-2854