Explaining the Illusion of Independent Agency in Imagined Persons with a Theory of Practice

By Jim Davies

The Version of Record of this manuscript has been published and is available in *Philosophical Psychology*, *2022*
https://www.tandfonline.com/doi/full/10.1080/09515089.2022.2043265

Abstract

Many mental phenomena involve thinking about people who do not exist. Imagined characters appear in planning, dreams, fantasizing, imaginary companions, bereavement hallucinations, auditory verbal hallucinations, and as characters created in fictional narratives by authors. Sometimes these imagined persons are felt to be completely under our control, as when one fantasizes about having a great time at a party. Other times, characters feel as though they are outside of our conscious control. Dream characters, for example, are experienced by dreamers as autonomous entities, and often do things that frighten and surprise dreamers. Some imagined persons, such as characters in fiction, start off under conscious control of the author, but over time, can appear to gain an illusion of independent agency. I propose an explanation for different autonomy attributions: characters are by default non-autonomous, unless their personalities are well-practiced. Characters become autonomous because modeling their thinking has become automatized, like many other well-practiced activities.

Address for correspondence: Prof. Jim Davies, Department of Cognitive Science, 2208 Dunton Tower, Carleton University, Ottawa, Ontario, Canada, K1S 5B6

Email: jim@jimdavies.org

1. Introduction

People can imagine make-believe situations, and often these situations involve characters. These imagined persons are sometimes felt to be under the imaginer's conscious control, as when, for instance, thinking about how one's self and spouse might allocate duties after work to make sure the children are picked up, the dog is walked, and dinner is cooked. One might suppose that this is a defining feature of imagination. However, it does not always obtain. There are times when imagined persons feel to be completely autonomous, outside of our conscious control. Dream characters are examples of this, as they often surprise and even frighten the dreamers who created them. Dreaming might not strike one as a prototypical example of imagination, but in this paper I will refer to an *imagined person* as any representation and mental simulation of a person in the mind of a person. This simulation might be brought about in many ways, including dreaming, hallucination, and voluntary fantasizing.

Some characters begin as feeling in conscious control (non-autonomous) but seem to become autonomous over time. A striking example of this is with some fiction writers, who claim that, after writing for a while, some of their major characters won't do what the authors need them to do for the plot, and getting them to behave requires negotiation.

Of course this feeling of imagined character autonomy is illusory—these characters do not have minds of their own, and the imaginer's mind is indeed controlling these characters, but the processing is subconscious, resulting in an *feeling* of autonomy that can be attributed to the imagined person, sometimes known as a *diminished sense of agency* (Bayne & Levy, 2006; Bergman-Hai, Kessler, & Soffer-Dudek, 2020) or the *illusion of independent agency* (the feeling of other-as-agent vs self-as-agent, Jones & Fernyhough, 2007). This is the feeling that one is engaging in actions that one did not voluntarily initiate. Strictly speaking, the characters are

never truly autonomous, but for simplicity I will refer to imagined persons as *autonomous* and *non-autonomous* as a shorthand for how imaginers experience them, rather than what they actually are. Finally, this illusion of independent agency can be experienced even if the imaginer doesn't *believe* that the agent is independent (Foxwell, Anderson-Day, Fernyhough & Woods, 2020). This contrast between our experience and the beliefs we endorse is also seen when one looks at an optical illusion that makes one line look longer than another: we can perceive that one line is longer yet simultaneously believe, after we understand that we are looking at an illusion, that they are the same length.

Why are some imagined persons autonomous, and others are not? This question has been largely ignored in the literature to date. There is only one theory that has tried to explain it: the idea that the illusion of independent agency might be the result of automatization was first introduced by Marjorie Taylor, Sara D. Hodges, and Adele Kohányi (2003). In brief, the idea is that our control of imagined persons becomes autonomous the same way our motor actions become autonomous with practice. However, this automatization theory was presented only in the context of fiction characters and imaginary companions.

In this article I will flesh out this idea and expand it to attempt to explain autonomy for *all* of the different kinds of imagined persons known to be simulated in people's minds. I will explore possible alternative explanations in a later section.

2. The Practice Theory

What I seek to explain is why some imagined characters feel to be autonomous and others non-autonomous, and others feel more autonomous over time. My theory is that autonomy can be explained using what we know about mental models and learning. Characters can feel

autonomous when the imaginer has a rich mental model of the character. This can be because it is based on a real or imagined person who is well known to the imaginer, or a *stock* character, such as "dangerous man," or "average kid."

Characters will feel non-autonomous when the imaginer has no rich mental representation, and as such requires the imaginer to use deliberative, conscious mental resources to control the character. Sometimes characters can start non-autonomous, but, through practice, the imaginer develops sufficiently detailed mental models such that the process of predicting the character's behavior becomes automatic and unconscious for the same reason that physical activities do through practice: automatization. At this point, the character will feel autonomous.

This happens because we become unaware of the processing involved with modeling the imagined person's behavior. When one first learns to type, they are aware of every finger and every press of the keyboard. Over time, one automatizes this process and can focus on other things, such as the content of what they want to write, while unconsciously carrying out those actions that consumed their conscious mind in the beginning. Phenomenologically, the keypress actions fade from consciousness, and it feels like your body knows what to do without your having to consciously monitor every finger's movement. The popular term "muscle memory" is instructive: you feel that your muscles know what they are doing, and you do not. When people attempt to describe how they do automatized activities, they often describe them incorrectly. For example, people use countersteering to turn a bicycle but are unaware, and will sometimes even deny, that they are doing this.

The psychological concept of *automatization* typically involves the attributes of uncontrollability, unconsciousness, efficiency, and automaticity. Some scholars hold that all four are required for something to be considered automatic, others that a particular case is more

automatic in relation to how many of these attributes are present, and some hold that autonomy is the only necessary condition (Bergman-Hai, Kessler, & Soffer-Dudek, 2020). In this paper, when I refer to automatization, I am referring to the attribute of autonomy: the feeling that the thought process is not under your conscious control.

Bringing it back to imagined persons, when one reasons about what a person would do, one will work through beliefs about the person—what they know, what their values and goals are, and so on. This use of a mental model of the person is a conscious process. But with practice, the mind automatizes reasoning with this particular mental model.[i] The lower-level symbolic manipulation fades from conscious focus. When enough of this happens, our imagined characters can behave in ways that surprise us. The surprise happens because our conscious minds are no longer privy to the reasoning that generated that behavior. This makes the imagined character feel autonomous.

In the following I will review the different kinds of imagined characters that have been documented and describe how well the theory explains and fails to explain differences in autonomy.

2.1. Imaginary Companions

Imaginary Companions, popularly known as imaginary friends, are make-believe, friendly characters, which seem to be created for the function of keeping the imaginer happy (Taylor, 1999). Most imaginary companion imaginers are children (41% of adults report having had them as children), but many hold on to their companions into late adolescence, and sometimes into adulthood. A recent survey of 1472 adults reported that 7% currently had imaginary companions (Fernyhough, Watson, Bernini, Moseley & Alderson-Day, 2019).

My theory predicts that when imaginary companions are first created, they would be under complete conscious control of the imaginer. Over time, the imaginer would get better and better practiced at imagining how the character would act and react, and as a result the process would become more and more automatized. Eventually, the imaginary companion would be completely autonomous.

There is good evidence that mature imaginary companions can be autonomous, though the amount of control the imaginer has varies (Davis, 2018). We know this from reports from imaginers, but also because the companions often misbehave, and act in ways counter to the imaginer's welfare. In a particularly striking example, a child's trip to a horse show was ruined because the imaginary companion had made other plans and couldn't accompany the child (Harris, 2000, p. 58).

What is unknown is whether or not the companions start out this way, as the research investigating this has not been done (Paige Davis, personal communication). One potential reason for this is that developmental psychologists' operational definition of an imaginary companion requires that they have been a part of the imaginer's experience for at least three months (Davis, 2018). As such, a child with an imagined playmate that was recently created might not qualify for inclusion in a study of imaginary companions at all. However, it might be that these proto-companions lead to full-fledged imaginary companions that psychologists might later identify as such. Future research on the development of imaginary companions over time could shed light on the issue of whether or not they start out as autonomous.

What we do know is that imaginary companions are played with a lot, supporting the notion that they are both autonomous and well-practiced, as my theory predicts. The current lack

of data on the development of imaginary companions means that the practice explanation for their autonomy is highly speculative.

2.2. Tulpas

A traditional Tibetan religious practice involves creating imagined entities called *tulpas*. Alexandra David-Neel's anthropological report on them (2012, cited in Juliani, N.D.) shows that these entities often appear as a result of great mental effort, but for someone very skilled in mental concentration, they might appear spontaneously. Apparently at some point in the exercise of creation, they become autonomous.

A recent phenomenon is that of adults deliberately creating imaginary companions, again, usually for the purpose of keeping the imaginer company. These companions, inspired by Tibetan Buddhism, also called "tulpas," are autonomous, and many imaginers, called "tulpamancers," report being able to successfully attenuate loneliness because of this imagined character. Unlike most unwanted hallucinated characters, interactions with tulpas are often pleasant (Martin, Thompson & Lancaster, 2020). A substantial percentage of them believe that tulpas are *actually* autonomous, and are the mind of some other being, although others believe that the automomy is illusory (Veissière, 2016). The scientific research on tulpamancy is still in its infancy, but there are long online discussions involving tulpamancers sharing advice and reporting experiences, as well as guidebooks for people who want to try to create one.

Similarly to imaginary companions, my theory predicts that tulpas would be non-autonomous upon creation, but over time would become more autonomous. This appears to be exactly the way it happens.

Tulpamancers often require months of deliberate visualization before their tulpas become autonomous and start to say things that surprise the tulpamancer. One guidebook states that it

takes about three months, or 200-500 hours, to form an autonomous tulpa (Thompson, 2014). We should be wary here of survivorship bias. Those who attempt to create a tulpa but fail to do so probably do not regularly post about their experiences on online forums. So we cannot say that practice reliably creates tulpas. We also have reports of tulpas being spontaneously created, without practice, which runs counter to the predictions of my theory. What we do know is that, for most people that have autonomous tulpas, they get them only after extended practice thinking about them.

Another striking feature of tulpas is that they cannot be easily dismissed. It takes a long time to get rid of one. Trying to dismiss a tulpa is like trying to stop being able to play the piano (Veissière, 2016). This is suggestive in that tulpamancy is compared to piano playing, a classic example of automatization in motor control. This analogy fails in an interesting way: although someone who can play the piano can choose to stop playing, someone with an unwanted tulpa, apparently, cannot help but continue to have the tulpa as a part of their experience without taking action to eliminate it.

In summary, what little we know about tulpamancy seems to generally support the proposed theory of autonomy in imagined characters, though the spontaneous generation of tulpas remains mysterious.

2.3. Fictional Characters in the Minds of Authors and Readers

When an author writes a novel, she must create and keep track of many characters. For all authors, the majority of their characters are non-autonomous. But for 61-92% of authors, some of their characters become autonomous (Taylor, Hodges, & Kohányi , 2003; Foxwell et al., 2020). These characters often refuse to act as the author would like them to in the story, resist the author's conscious desires for what the story should be like (sometimes requiring negotiation),

and give the authors advice about how to live their lives (Watkins, 2000). Here is a vivid example: Alice Walker, author of *The Color Purple*, claims to have "lived" with two of her main characters for an entire year, chatting with her and telling her their backstories (Walker, 1972). Another author said: "They develop their own narratives, rapidly accumulating their own histories and anecdotes - if unchecked, I have had to kill off characters to stop a story digressing out of sight" (Foxwell, 2018).

My theory predicts that fictional characters would only become autonomous after the authors have spent considerable time thinking about their mental states and behaviors. This appears to be supported by the available evidence.

First, only major characters seem to become autonomous. Minor characters, such as a restaurant server who appears only in a single scene, is very unlikely to refuse to behave. Second, even major characters seem to take time to become autonomous—perhaps halfway through the book. As one respondent says, "At the start of a book, I write more slowly and it's much more painful as I'm still trying to 'tune' the characters in. As they start speaking in my head, it becomes easier and the writing speeds up and becomes more fun" (respondent R157 in Foxwell et al., 2020). Some authors have reported that characters become autonomous after writing thousands of words. The average is 30,000 words, or about 100 pages (Foxwell, 2018). A part of my theory is that autonomous characters who are not well-learned will be based on stock character ideas in memory. In support of this, one author reported that if the illusion of independent agency does not happen, the novel suffers by "relying on archetypes" (Foxwell, 2018). A related phenomena is that actors sometimes feel that the characters they are playing have autonomy (Brown, 2019). However, there is no data on how long it takes for this to happen.

My theory predicts that stock characters will be autonomous, but this does not appear to be the case for fiction writers, who may have non-major characters that one might describe as archetypical or stock, but are not autonomous.

It's possible that some autonomous fictional characters appear to writers without practice, as the rare tulpa does. If this happens, it is counterevidence to the automatization theory. I am aware of no reports of this happening in the case of writers, however.

Even readers can experience book characters after the book is put down. This happens to nearly a fifth of them (Alderson-Day, Bernini, & Fernyhough, 2017). Understanding fiction requires substantial cognitive work on the part of the reader—using mental models of how the world works and, relevant to the present discussion, use of theory of mind to understand the characters. Like authors, readers develop schemas to understand each character. On my explanation, repeated use of a schema can lead to automatization and the illusion of independent agency. Like authors, readers experience what characters would do or say without using "deliberative empathizing or reasoning about the characters' mental states" (Foxwell et al., 2020), suggesting automatization.

The similarity of autonomous fiction characters and autonomous imaginary companions has not escaped other researchers. Taylor, Hodges, and Kohányi explicitly endorse a practice/automatization theory of how autonomy develops (2003).

2.4. Dream Characters

Characters encountered in dreams are almost always autonomous. And, like imaginary companions, they can surprise and frighten the dreamer/imaginer.

This is even true for characters in lucid dreams. Lucid dreams are dreams in which the dreamer is aware that they are dreaming. About 80% of people will experience one at some point

in their lives (Voss, 2014). Typically, this makes dreams more vivid, allows dreamers more control over their dream avatars, and have some control over the dream environment. However, for most lucid dream experiences, dream characters in the dream world are autonomous—so much so that psychologists have done experiments in which lucid dreamers are instructed to ask their dream characters questions, to see what they're like and how they behave (Stumbrys, Erlacher & Schmidt, 2011).

Because all dream characters are autonomous, and it is generally impossible to control dream character behavior with conscious processing, my theory predicts that all dream characters would all be based on well-learned models. These models could be stock characters, or based on real people the dreamer knows well, or fictional characters that are well-practiced. In particular, my theory predicts that no dream character would ever be psychologically complex but *not* based on one of these three existing kinds of models.

Unfortunately, no empirical research to date directly bears on this question. We do know that it is quite common to be chased by a dangerous male in a dream, suggesting that "dangerous male" is one of the standard stock characters people have in their mental inventory. Why would stock characters also be autonomous? Because a simple character requires much less practice to automatize than a complex one, just as a simple physical task (like pressing a button when you see a light) takes less practice to automatize than a complex one (like tango dancing). It's possible that stock characters such as "dangerous man" are thought about so frequently, in fantasy, in storytelling, and so on, that its simple mental model becomes automatized somewhat easily.

When one dreams of someone they know only a little bit, my theory would predict that this person would take on the psychological profile of some stock character.

Why all dream characters (with rare exceptions) are autonomous is still mysterious. Why, when dreaming of an acquaintance we do not know well, are we out of conscious control of them? I can only speculate: a major class of theories of dream function hold that dreams are practice for the real world (perhaps threatening or social situations, see Revonsuo 2000 and Revonsuo, Tuominen & Valli, 2015, respectively), and if we were to control some part of it beyond our dream avatar (ourselves in the dream), the practice would be less effective, because it would be a salient clue that we were dreaming, potentially causing us to take the dream content less seriously, and thereby not acting realistically—which makes for better practice.

2.5. Hallucinated Characters

Hallucinations can range in complexity, from simple lights or colors (as in migraine auras), to full-world dream-like experiences that do not incorporate concurrent sensory experience (as in hallucinations caused by temporal lobe epilepsy or taking the drug DMT). More complex hallucinations can involve human characters, or other creatures who are psychologically human. I will refer to all of these as "hallucinated people." Autonomy is a common characteristic of hallucinated people.

The most common type of hallucinated person is a voice in the head. A common symptom of schizophrenia (60%-74% of persons with schizophrenia have them, though non-clinical populations sometimes do as well; Slade & Bentall, 1988; Hill & Lindon 2013, p. 21), heard voices are almost universally autonomous (though the level of autonomy varies, Wilkinson & Bell, 2016). Hallucinated characters tend to be shallow, and not as deep as literary characters. (Watkins, 2000, p. 115). 80% of hallucinated voices are experienced as independent agents (Corstens & Longden, 2013). About 47.5% of voice-hearers can engage in interactive conversations with the voices (Alderson-Day et al., 2021; Leudar, Thomas, McNally & Glinski,

1997). Often, they are perceived to be coming from the external environment. That is, a voice-hearer might hear a voice, and infer that it is coming from the fire escape or the next room. Sometimes the voice-hearer will figure out that the voices are not coming from the external physical environment and might then infer that they are coming from his or her own mind, or perhaps from some other source, such as aliens or government agencies. But the phenomenal quality of the voice can still sound like it's coming from somewhere—in the mind it has auditory localization (Woods, Jones, Alderson-Day, Callard & Fernyhough 2015). There are many non-clinical cases of voice hearers, as well. People claiming to be clairaudient psychics might experience heard voices every day. Although they have more control than clinical populations over the onset and offset of the voices, they do not feel to be in control of what the voices say, which likely encourages their supernatural beliefs that the voices are those of spirits or ghosts (Powers, Kelley, & Corlett, 2017).

One common kind of hallucinated person comes from bereavement hallucinations, which are when people who lose a loved one hallucinate the deceased shortly after death. They might hear the voice of the deceased loved one, or perhaps see them in a chair. These persons are reported to be autonomous (Sacks, 2012, pp. 253—254; Rees, 1971; Keen, Murray, & Payne, 2013), as are most hallucinated persons. Social isolation can also lead to hallucinated people (Hoffman, 2007).

Charles-Bonnet syndrome also can involve hallucinated persons, but they are always purely visual, and do not interact with the hallucinator (Burke, 2002). A Charles-Bonnet syndrome patient might hallucinate people walking through the room, often in strange costumes. Charles-Bonnet syndrome is primarily a perceptual problem associated with loss of vision. Perhaps because other mental faculties are untouched, these patients quickly gain insight, and

recognize the hallucinations for what they are. But they never seem to be able to control what they hallucinated characters do.

From a scientific perspective, characters in experiences described in terms of religious revelation experiences are classified as hallucinated persons. Gods and divine messengers (*guiding presences*) are autonomous, by definition—any non-autonomous god-like character would probably not be interpreted as a part of a revelation. A non-autonomous consideration of a divine being, as when someone muses on what a God's opinion on an issue might be, would be classified by the imaginer as a case of daydreaming, or hypothetical reasoning, instead of religious revelation. In practice, hallucinated presences by religious people are subject to interpretation: is this particular experience a run-of-the-mill hallucination, or an actual revelation?

Ann Taves describes three cases of guided presences in history, and in all cases the guided presence do not appear to emerge, fully-formed, right away (Taves, 2016). For Joseph Smith, the founder of Mormonism, Helen Schucman, who "scribed" the book *A Course in Miracles* (Schucman & Thetford, 2009) based on what her guided presence told her, and Carl Jung's *The Red Book* (2009) all involved meditation, active imagination, and lengthy interpretation with the help of others preceding the illusion of independent agency. For Schucman, for example, the voice gradually emerged over time, becoming more and more insistent that she write down what the presence said and share it with the world. In all three of these cases, the presences became autonomous. Jung and Schucman were "reluctant prophets" that were regularly chastised by their guiding presences for not doing what they were told. The fact that goals of the presence conflict with the hallucinator's goals is often used as evidence that the presence does not come from the hallucinator's mind (Taves, 2018). For our purposes, it

suggests the experience of autonomy. When *consciously* thinking about God's moral opinions, for example, people tend to report that God's opinions are the same as their own (Epley, Converse, Delbosc, Monteleone & Cacioppo, 2009). But when experiencing God or a god as a hallucinated character, their opinions can differ.

Though these presences seem to change over time, from Taves's research it is not clear that their autonomy changes over time. However, Tanya Lurhman's anthropological research of evangelical Christians shows that training is usually necessary to create an autonomous experience of God (Lurhman, Nusbaum & Thisted, 2010). Some Christian communities train participants to practice imagining God (walking with him, talking to him, and so on) with the goal of the experience becoming autonomous. When this happens, the imaginer comes to believe that they have established a connection to God, and that the character they are experiencing is actually God.

For those hallucinated characters (religious or otherwise) that are autonomous from the start, my theory has the same predictions of them as it does with dream characters: that all heard voices are either stock characters, or models of well-practiced people, be they real or fictional. Although there is no systematic study of whether hallucinated characters grow in complexity over time, one particularly thoughtful subject of hallucinated voices, Gregory Shankland (personal communication, 2020), reports that the several voices in his head have grown in complexity over the many years he's heard them, and that although the different voices claim to be different characters in the different narratives they engage in, each voice has a distinct personality that comes through in all of the characters—much like you might detect common mannerisms of an actor playing different roles in various films.[ii]

My theory would predict that as a hallucinator gained experience with a particular hallucinated character (some have persistent personalities for years), the character would get richer and more complex. Data that would speak to this have not been collected, though Shankland's anecdote is suggestive. It is possible that viewing hallucinated characters as a single class of experiences is unwise, as there may be multiple brain mechanisms that give rise to them, such as memory intrusion, misattributed inner speech, or random activation of the auditory cortex (Wilkinson & Bell, 2016).

2.6. Daydream Characters

A daydream might be a fantasy, such as fantasizing about winning an award, or a relaxation technique, such as imagining being on the beach. A daydream might involve anxiety, such as visualizing terrible things that might happen, such as your children being hit by a car. Daydreaming might also involve simple planning, such as when you imagine the best route through the city to run all of your errands.

Daydreams often, but not always, involve imagined persons. Sexual fantasies often involve a sexual partner, planning often involves interactions with imagined characters, and so on. My theory predicts that daydream characters will be autonomous if they are well-known or stock characters and that any complex character that is not well-practiced will be non-autonomous.

There is no empirical research describing the autonomy of daydream characters. Commonsense experience strongly suggests, however, that daydream characters *do not* appear to be well-described by my theory, because all daydream characters appear to be non-autonomous. For example, suppose Julie is planning a surprise birthday party for her husband of ten years, Rick. She imagines how Rick might react to the surprise, where he might be most comfortable

having the party, and so on. She is drawing on her mental model of Rick to predict his behavior, but in the daydream Rick does not really surprise her, in the same way a dream character might. However, her mental model of Rick is very rich. She has interacted with him every day for over a decade. My theory suggests that her model of Rick would be so well-practiced that every time she imagined Rick, he would be completely autonomous. But that's not what happens.

If Julie's imagining planning who will do what after work, she might imagine that Rick will come home and walk the dog, while she picks up the kids and goes grocery shopping. As she imagines this, her imagined Rick won't object and instead drive somewhere to play pool, in spite of her wanting to imagine him walking the dog. Although she might know that Rick might not be keen to walk the dog, it's not as though the imagined Rick would "misbehave" in the way some authors report some of their autonomous characters do. Further evidence that her daydream of Rick is not autonomous is that she can easily imagine Rick doing something out of character, perhaps even very vividly. If the imagined Rick were autonomous, this would be difficult, as many fiction authors have reported experiencing.

Further, when she has a dream of Rick, the dream version of her husband *is* autonomous. Here we have the same imagined character acting autonomously in a dream, but non-autonomously in a daydream, regardless of the practice the imaginer has experienced with thinking about the character. So, anecdotally, daydream characters seem to defy the theory presented in this paper.

3. Alternative Theories

The theory of practice is the only theory published that has explicitly been proposed to explain the illusion of independent agency. In this paper I have expanded on this idea and shown

evidence for and against it. But what other explanations might one propose? Here I will discuss two possibilities.

The Efference Copy Theory. One explanation for the feeling of other-as-agent comes from the hallucination literature. According to one theory, when a motor action is planned, the motor command is sent to the muscles, and at the same time a version of this command is sent to the perceptual systems to anticipate the expected sensory experience. For example, a part of Julie's mind might decide to slap her hand on the table. As such, her perceptual system is primed to expect the feeling of the arm moving and the hard smack of something on the palm. This message to the perceptual system is called the *efference copy*. We pay much less attention to perceptions that match the content of the efference copy. When there is a problem with this *forward modeling* system, then our perceptions do not match expectation (Ford & Hofman, 2013, p. 362). For example, if the efference copy does not arrive, then the perceptual system will be (metaphorically) surprised at the feeling of motion and the smacking. This is theorized to explain alien hand syndrome and other kinds of other-control delusions and hallucinations.

Jones and Fernyhough (2007) apply this to auditory verbal hallucinations. Building on the idea that inner speech is an action, a failure of the forward modeling system might result in a similar lack of feeling of self-as-agent, and possibly lead to an inference or perception that the voice heard comes from elsewhere.

Although this theory could be right for some instances of imagined persons, particularly those generated as a result of mental illness, it leaves some important things we know about imagined persons unexplained. For example, it says nothing about how imagined persons can seem to get more autonomous over time. Why, for instance, would writing about a fictional

character for 30,000 words, or a tulpamancer's three months of visualization practice, result in a failure of the forward modelling system when the character becomes autonomous?

Another problem is that any given individual might have some instances of imagined persons that are autonomous and others that are not. For example, just because a fiction writer has some autonomous characters does not mean that *all* the characters are autonomous. On the contrary, as we have seen, typically it is only the main characters who become autonomous. Similarly, a person might have a few imaginary companions, but this does not mean that the illusion of independent agency is present in every person they might imagine. This rules out the possibility that the problem with the efference copy is a global trait. We might modify this view such that an efference copy is not sent *for some imagined persons but not others*, but then we would need some description of when this would happen and when it wouldn't—and a theory of practice might help here.

Further development of this theory might explain how forward modelling is not functioning in dreams, or why it does not function properly for non-clinical populations like those with imaginary companions, but right now the theory is silent on these issues.

The Dissociative Absorption Theory. Although it has not been specifically proposed as a broad explanation for the illusion of independent agency, we can explore the idea that having the trait of having higher-than-average levels of *dissociative absorption* as explaining autonomy in one's imagined persons (Bergman-Hai, Kessler, & Soffer-Dudek, 2020). Dissociative absorption is when one's awareness becomes completely engrossed in an internal or external stimulus, at the expense of being oblivious to one's physical environment. One's general level of dissociative absorption (that is, as a trait) is measured by the Dissociative Experiences Scale (DES; Carlson et al., 1993). Low scores on the DES are non-pathological, but high scores are associated with

dissociation disorders, and a very dissociation might reflect the presence of Dissociative Identity Disorder (Simeon et al., 2009). A similar construct is simply known as *absorption*, and is measured with a different test, the Tellegen Absorption Scale (TAS; Tellegen & Atkinson, 1974.) DES and TAS scores are moderately correlated (Levin & Spei, 2004).

How might dissociative absorption explain the illusion of independent agency? Dissociative absorption is characterized by an intense attention and consciousness of one aspect of internal or external stimulus, and a dramatically reduced attention to everything else. If someone is prone to dissociative absorption, then if they are not focused on the (simulated) intentions of an imagined person then they will not be conscious of them, resulting in an illusion of independent agency. If dissociative absorption is the reason for the illusion, then we should see higher levels of it (as measured on the DES) in those who more often experience the illusion.

There is evidence to support this position. An empirical study of populations of fiction writers showed a positive correlation between dissociative absorption and autonomous characters in their writing (Taylor, Hodges, & Kohányi , 2003). Children who have imaginary companions also score high on absorption (Kidd, Rogers, & Rogers, 2010).

People with high levels of dissociative absorption also are worse at remembering content they have generated themselves, but not content they take in through their senses; these participants also report a subjective feeling of a diminished sense of agency in general (Bergman-Hai, Kessler, & Soffer-Dudek, 2020). Studies of patients with dissociation-related disorders show that they have problems with source attribution (Chiu et al, 2016.) Note that these studies look at one's sense of agency for self-generated content in general, not specifically that of imagined persons. Imagined persons are one example of self-generated content.

The dissociative absorption theory I have sketched out here suggests that it is one's level of dissociative absorption that predicts autonomy in imagined persons. It would predict that the autonomy of an imagined person would be just one example of a lack of source monitoring insight in general. That is, a person high in the trait of dissociative absorption would have trouble attributing their own thoughts and actions to themselves, and their imagining of people is just one example of this.

This theory alone has trouble explaining why there is variance in the autonomy of imagined person in the same individual. One person might have autonomous imagined persons in dreams, but full control of their imagined persons in daydreams. A person might have an autonomous imaginary companion, but not every person they imagine has this trait. It also does not explain why fiction writers would only have a few autonomous characters, and why this tends to happen only after writing about them for many chapters.

I do not want to suggest that these alternative explanations are incompatible with each other or with the theory of practice I have presented in this paper. It could very well be the case that an imagined person will tend to become autonomous for an individual with practice and also be true that this process is facilitated and accelerated by high dissociative absorption or problems with efference copies. A multi-factor explanation might be correct.

4. Conclusion

My theory is that when imagined characters are complex but not well-practiced, they will always be non-autonomous. That is, they will feel to be under the control of the imaginer. Imagined characters who feel autonomous are so because they are well-understood or well-practiced. That

is, the mental model for that character is so rich that it can be used unconsciously. These might be "stock" characters that are reasoned about frequently, or well-practiced complex characters, such as some major characters in author's stories, or people one knows well.

From what we know of imagined characters, in their various forms, the theory presented here predicts some of what we observe in human experience, but not everything.

The theory predicts well some authors' experiences of major characters becoming autonomous only after many hours of rehearsal. This is also true of the relationship between tulpamancers and their tulpas. This also might be true of imaginary companions, but our lack of data on the autonomy of proto-imaginary companions means we just do not know yet. We simply do not have data to support or falsify the theory for imaginary companions.

For dream characters and hallucinated characters, we also suffer from a lack of data in terms of whether they are all either stock or well-practiced, as the theory predicts. If people can dream of poorly-practiced, complex characters, it would be counterevidence to this theory.

Daydreams pose a clear challenge to the theory, because in typical daydreams (anecdotally), all characters are non-autonomous, including ones that are well-practiced, such as those of people close to us and stock characters. The same imagined character will be non-autonomous in a daydream but autonomous in a dream, regardless of how complex the mental model is for that character.

A further problem is that there are some major differences between how automaticity works for motor action and imagined persons. An automatized action, such as driving, is guided by conscious goals, but the lower-level muscle actions are automatic. They are rather inflexible, and predictable. When a mind creates dream character behavior, there are none of these features. This

inconsistency with other forms of automatization makes the practice theory more suspect. My theory has no explanation for these exceptions.

In conclusion, it appears that ideas of mental models of people, and automatization ideas from what we know about learning, are a good start to a theory of why some imagined characters are autonomous and some are not. But much data are missing, and some data appear to contradict it, which means the practice theory is, at best, only a part of the answer.

Some intriguing questions remain unanswered: are imaginary companions autonomous right away? How long does it take for imaginary companions to become autonomous? Does this length of time match that of the experiences of tulpamancers and some authors? Are hallucinated and dream characters always well-known or stock? Why are some well-practiced characters (such as our loved ones) autonomous in dreams but non-autonomous in daydreams and fantasies?

Some scholars break the phenomenology of agency into components, such as mental causation, authorship, and effort (Bayne & Levy, 2006). Mental causation is a feeling that something happened in your mind to cause the action or thought. Authorship is the feeling that you, voluntarily, caused the action or thought, and effort reflects the feeling that one had to do some mental work to make it happen. Reports on the illusion of independent agency conflate these constituent feelings. But do they always go together? When someone reports that they feel out of control of a character, it might, on some occasions, mean that they feel authorship but no mental causation where such a feeling was expected. Empirical investigations of these constituent feelings in dreaming, imaginary companions, and so on might allow for a more nuanced understanding of the phenomenon.

Acknowledgements

Several anonymous reviewers for this paper were very helpful in fleshing out some of the nuances of this issue.

References

Alderson-Day, B., Bernini, M., & Fernyhough, C. (2017). Uncharted features and dynamics of reading: Voices, characters, and crossing of experiences. *Consciousness and Cognition*, *49*, 98-109.

Alderson-Day, B., Woods, A., Moseley, P., Common, S., Deamer, F., Dodgson, G., & Fernyhough, C. (2021). Voice-hearing and personification: characterizing social qualities of auditory verbal hallucinations in early psychosis. *Schizophrenia bulletin*, *47*(1), 228-236.

Bayne, T., & Levy, N. (2006). The feeling of doing: Deconstructing the phenomenology of agency. In S. Sebanz, & W. Prinz (Eds.). *Disorders of volition* (pp. 49–68). Cambridge, MA: MIT Press.

Bregman-Hai, N., Kessler, Y., & Soffer-Dudek, N. (2020). Who wrote that? Automaticity and reduced sense of agency in individuals prone to dissociative absorption. *Consciousness and Cognition*, *78*, 102861.

Brown, G. H. (2019). *Blurred Lines Between Role and Reality: A Phenomenological Study of Acting*. Doctoral dissertation, Antioch University.

Burke, W. (2002). The neural basis of Charles Bonnet hallucinations: A hypothesis. *Journal of Neurology, Neurosurgery, and Psychiatry,* 73, 535—541.

Carlson, E. B., Putnam, F. W., Ross, C. A., Torem, M., Coons, P., Dill, D. L., ... Braun, B. G. (1993). Validity of the dissociative experiences scale in screening for multiple personality disorder: A multicenter study. *American Journal of Psychiatry*, 150(7), 1030–1036.

Chiu, C. De, Tseng, M. C. M., Chien, Y. L., Liao, S. C., Liu, C. M., Yeh, Y. Y., & Hwu, H. G. (2016). Misattributing the source of self-generated representations related to dissociative and psychotic symptoms. *Frontiers in Psychology*, 7, 1–10.

Corstens, D., & Longden, E. (2013). The origins of voices: links between life history and voice hearing in a survey of 100 cases. *Psychosis*, *5*(3), 270-285.

David-Neel, M. A. (2012). *Magic and mystery in Tibet*. New York: Courier Corporation.

Davis, P. (2018). Adult report of imaginary companion play in childhood and its relation to concurrent prodromal symptom report. Talk at *Personification Across Disciplines 2018*, Durham, UK.

Epley, N., Converse, B. A., Delbosc, A., Monteleone, G. A., & Cacioppo, J. T. (2009). Believer's estimates of God's beliefs are more egocentric than estimates of other people's beliefs. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21533-21538.

Fernyhough, C., Watson, A., Bernini, M., Moseley, P., & Alderson-Day, B. (2019). Imaginary companions, inner speech, and auditory verbal hallucinations: What are the relations?. *Frontiers in Psychology*, 10, 1665.

Ford, J.M. & Hofman, R.E. (2013). Functional brain imaging of auditory hallucinations: From self-monitoring deficits to co-opted neural resources. In R. Jardri, A. Cachia, P. Thomas,

& D. Pins (Eds.) *The neuroscience of hallucinations*. (pp. 359--373). New York, NY: Springer.
Foxwell, J. (2018). 'I won't be involved with this fictional plot:' Characters' agency and authors' intentions. Talk at *Personification Across Disciplines*, Durham, UK.
Foxwell, J., Alderson-Day, B., Fernyhough, C., & Woods, A. (2020). 'I've learned I need to treat my characters like people': Varieties of agency and interaction in Writers' experiences of their Characters' Voices. *Consciousness and Cognition*, *79*, 102901.
Harris, P. L. (2000). *The work of the imagination*. Oxford, UK: Blackwell Publishers Ltd.
Hill, K. & Linden, D.E.J. (2013). Hallucinatory experiences in non-clinical populations. In R. Jardri, A. Cachia, P. Thomas, & D. Pins (Eds.) *The neuroscience of hallucinations.* (pp. 21-41). New York, NY: Springer.
Hoffman, R. E. (2007). A social deafferentation hypothesis for induction of active schizophrenia. *Schizophrenia bulletin*, *33*(5), 1066-1070.
Jones, S. R., & Fernyhough, C. (2007). Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition*, *16*(2), 391-399.
Jung, C. (2009). *The Red Book.* New York: W.W. Norton.
Juliani, A. (N.D.). On tulpas. *Unpublished Manuscript.* Retrieved May 12, 2020 from http://www.academia.edu/download/41047885/On_Tulpas.pdf
Keen, C., Murray, C., & Payne, S. (2013). Sensing the presence of the deceased: A narrative review. *Mental Health, Religion & Culture*, *16*(4), 384-402.
Kidd, E., Rogers, P., & Rogers, C. (2010). The personality correlates of adults who had imaginary companions in childhood. *Psychological Reports*, 107(1), 163-172.
Leudar, I., Thomas, P., McNally, D., & Glinski, A. (1997). What voices can do with words: pragmatics of verbal hallucinations. *Psychological medicine*, *27*(4), 885-898.
Levin, R., & Spei, E. (2004). Relationship of purported measures of pathological and nonpathological dissociation to self-reported psychological distress and fantasy immersion. *Assessment*, 11(2), 160–168.
Luhrmann, T., Nusbaum, H. & Thisted, R. (2010). The Absorption Hypothesis: Learning to Hear God in Evangelical Christianity. *American Anthropologist*, Vol. 112/1, 2010, 66-78.
Martin, A., Thompson, B., & Lancaster, S. (2020, May 11). Personality characteristics of tulpamancers and their tulpas. *PsyArXiv* https://doi.org/10.31234/osf.io/5t3xk
Powers III, A. R., Kelley, M. S., & Corlett, P. R. (2017). Varieties of voice-hearing: psychics and the psychosis continuum. *Schizophrenia bulletin*, *43*(1), 84-98.
Rees, D. W. (1971). The hallucinations of widowhood. *British Medical Journal*, 4, 37—41.
Revonsuo, A. (2000). The reinterpretation of dreams: An evolutionary hypothesis of the function of dreaming. *Behavioral and brain sciences*, *23*(06), 877--901. Page 891.
Revonsuo, A., Tuominen, J., & Valli, K. (2015). The avatars in the machine: Dreaming as a simulation of social reality. In Open MIND. Open MIND. Frankfurt am Main: MIND Group.
Sacks, O. (2012). *Hallucinations*. New York: Vintage.
Schucman, H. & Thetford, W.T. (Ed.) (2009). *A course in miracles.* New York, NY: Course in Miracles Society.
Simeon, D., Giesbrecht, T., Knutelska, M., Smith, R. J., & Smith, L. M. (2009). Alexithymia, absorption, and cognitive failures in depersonalization disorder: A comparison to posttraumatic stress disorder and healthy volunteers. *The Journal of Nervous and Mental Disease*, 197(7), 492–498.

Slade, P., & Bentall, R. (1988). Sensory deception: Towards a scientific analysis of hallucinations. London: Croom Helm.

Stumbrys, T., Erlacher, D., & Schmidt, S. (2011). Lucid dream mathematics: An explorative online study of arithmetic abilities of dream characters. *International Journal of dream research*, *4*(1), 35-40.

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, 83(3), 268–277.

Taves, A. (2016). *Revelatory events: Three case studies of the emergence of new spiritual paths*. Princeton, NJ: Princeton University Press.

Taves, A. (2018). Guiding presences and the emergence of new revelation. Talk at *Personification Across Disciplines* conference (PAD-2018). September 17-19, Durham, UK. https://vimeo.com/296826197

Taylor, M. (1999). *Imaginary companions and the children who create them*. New York: Oxford University Press.

Taylor, M., Hodges, S. D., & Kohányi, A. (2003). The illusion of independent agency: Do adult fiction writers experience their characters as having minds of their own? *Imagination, Cognition and Personality, 22*, 361-368.

Thompson, N. (2014). The internet's newest subculture is all about creating imaginary friends. *Vice.* Retrieved February 4, 2020 from https://www.vice.com/en_ca/article/exmqzz/tulpamancy-internet-subculture-892 See also https://www.tulpa.info/archive/tulpa-guide-for-dummies/

Veissière, S. (2016) Varieties of tulpa experiences: The hypnotic nature of human sociality, personhood, and interphenomenality. In A. Raz and M. (Eds). *Hypnosis and meditation: Towards an integrative science of conscious planes*. Oxford, UK: Oxford University Press. Pp. 55—76.

Voss, U. (2014). Unlocking the lucid dream. *Scientific American*, winter, special issue on creativity. 66—69.

Walker, A. (1972). In search of our mothers' gardens. *Worlds of difference: inequality in the aging experience*, 48-53.

Watkins, M. (2000). *Invisible guests: The development of imaginal dialogues.* Woodstock, CN: Spring Press.

Wilkinson, S. & Bell, V. (2016). The representation of agents in auditory verbal hallucinations. *Mind & Language*, *31*(1), 104-126.

Woods, A., Jones, N., Alderson-Day, B., Callard, F., & Fernyhough, C. (2015). Experiences of hearing voices: analysis of a novel phenomenological survey. *The Lancet Psychiatry*, *2*(4), 323-331.

[i] Foxwell et al. (2020) refer to this as "a schema for the character's perspective" and Alderson-Day et al. (2017) refer to this a "personality model."

[ii] You can learn more about Gregory Shankland at https://onfrontiers.com/profile/gshankland