

## Chapter 5

# The Justification Hypothesis

On June 11, 1963, an elderly Buddhist monk named Thich Quang Duc calmly walked to the center of a circle of protesters, sat down on a cushion, and meditated in the lotus position as he was doused by his religious brothers in a specially concocted mix of gasoline and diesel fuel. He then proceeded to light a match. Although he reportedly grimaced in agony, he did not move, scream, or cry out as his body was incinerated.

What could possibly account for such a dramatic and seemingly unnatural human act? Common psychological constructs such as schedules of reinforcement, personality traits, neural networks, unconscious conflicts, social roles, and evolved predispositions do not appear to provide a framework that would allow for the meaningful interpretation of such behavior. It would seem then that journalists, poets, and philosophers would do as well or better than psychologists in offering a viable account of Quang Duc's actions.

This chapter presents a third piece of the unified theory, the Justification Hypothesis. Whereas Behavioral Investment Theory provides a framework that allows for the understanding of how human behavior is continuous with other animals, the Justification Hypothesis provides the framework for understanding what makes people such unique animals. It is an idea that casts the relationship among language, human self-consciousness, and the evolution of Culture in a new light by interpreting both human self-consciousness and Culture as justification systems. As described in [Chapter 1](#), justification systems are the interlocking networks of language-based beliefs and values that function to legitimize a particular version of reality or worldview.

Viewing the behavior of Quang Duc and his fellow monks through the lens of the Justification Hypothesis, the following question emerges: What was the justification for his actions at both individual and societal levels? One inference that can readily be made is that the coordinated behaviors of the group of monks suggest the presence of a shared justification narrative; that is, everyone involved in the protest shared the same script. Moreover, the strikingly destructive and agonizing nature of the act implies that it could be legitimized by only the strongest of justification pressures.

How does the understanding afforded by the Justification Hypothesis fare in explaining Quang Duc's actions? Before his final dramatic act, Quang Duc left a

letter, now known as the Letter of Heart Blood, which justified his sacrifice as an attempt to demonstrate to the world the magnitude of injustices that were being perpetrated on the Buddhist community by a repressive regime. His plan worked well, in part because the act was captured by a Western photographer and entered the world's consciousness. Many nations consequently brought pressure on the South Vietnamese government to soften its attitude toward the traditional religion, and ultimately it complied. Quang Duc's justification for his self-sacrifice can thus be situated within a larger constellation of cultural justification narratives and—according to the Justification Hypothesis—it is these very narratives that provide the key to interpreting a wide range of human behaviors.

It is, of course, one thing to acknowledge that human beings often feel obliged to justify their actions. It is quite another to suggest that the phenomenon of justification represents a path to a comprehensive understanding of self-consciousness and the evolution of human culture. In this chapter, I show how the Justification Hypothesis provides a clear evolutionary explanatory framework that explicitly links the design features of human self-consciousness with a novel adaptive problem faced by our ancestors. I then introduce a tripartite model of human consciousness based on the insights afforded by the Justification Hypothesis that consolidates many different perspectives and allows for a readily understandable frame from which to view the major domains of human conscious experience. Then, in much the same fashion that Behavioral Investment Theory builds bridges between many disparate domains of thought connecting cognition, brain, and behavior, and the Influence Matrix connects many domains of research on human relational processes, I explain how the lens of the Justification Hypothesis assimilates and connects many lines of research in cognitive, social, developmental psychology, and neuropsychology. Finally, I explore how the Justification Hypothesis links individual level analyses with cultural level processes, providing a framework for understanding the evolution of culture and building a bridge between human psychology and the social sciences such as sociology and anthropology.

## **Linking Human Self-Consciousness to the Adaptive Problem of Social Justification**

Although there have been a number of proposals about the rise of self-consciousness and the rapid evolution of human culture, it nevertheless remains the case that the human capacities for self-consciousness and culture remain one of the great mysteries in the natural sciences. Consider that the co-discoverer of natural selection, Alfred Russell Wallace, was so perplexed about the human capacity for self-reflective thought that he thought it evidence for the divine. Almost a century later, the famed evolutionary biologist George Williams (1966) wrote, “I cannot readily accept the idea that advanced mental capabilities have ever been directly favored by selection” (p. 14). Likewise, Lumsden and Wilson (1983) characterized it as the “greatest missing link” in evolutionary theory, a characterization more recently echoed by Terrace and Metcalf (2005).

The Justification Hypothesis links a unique, biologically adaptive problem faced by our ancestors to the design features of human self-consciousness. Put in the jargon of evolutionary biology, the Justification Hypothesis is an exercise in reverse engineering. Capitalizing on Darwin's fundamental insight that the complex design seen in organisms is a product of natural selection, a reverse engineer matches organism design features to problems in the ancestral environment. To offer an example, let's say one was an alien visitor and wanted to understand why humans have outer ears. Using a reverse engineering approach, one might hypothesize, for example, that outer ears function to attract mates, much like a peacock's tail. That would lead to a number of testable implications. If ears had a primary role in mating rituals, then ears should be displayed more prominently during such times, and ear size and shape should correlate with sexual attractiveness, and ears should be center pieces of adornment and so forth. Of course, this hypothesis does not hold very well, and because the predictions don't align with observations one can reject the idea that the outer ear functions primarily to attract mates. Another hypothesis—that the outer ear functions to funnel sound waves into the inner ear thus enhancing hearing capacities—is a much better idea. Hearing capacity is indeed enhanced by the outer ear, and the shape of the ear is designed precisely in a manner that coordinates the flow of sound waves into the ear canal.

So reverse engineering is like a theoretical lock-and-key matching process, whereby one can think of the characteristic of the organism as the "key" and the adaptive problem it evolved to solve as the "lock." The shape of the outer ear partially "unlocks" the problem of getting sound waves into the inner ear. Now we can use this metaphor and turn our attention to human self-consciousness. When dealing with human self-consciousness there are problems with both the lock and the key. First, there is enormous difficulty conceptually specifying the "key"—that is, answering with clarity, what exactly the self-consciousness system is (recall some of the arguments visited in [Chapter 2](#) regarding problems in the philosophy of mind and consciousness). Second, as the previous quote from George Williams suggests, there have been few compelling accounts of biological problems that the self-consciousness system would have evolved to solve. The power in the Justification Hypothesis is that, for the first time, it affords us a clear way to conceptualize both the "key"—that is, defining the key design features of the human self-consciousness system, and a clear way to conceptualize the "lock"—the novel adaptive problem the system evolved to solve.

So according to the Justification Hypothesis, what is the self-consciousness system? In a nutshell, it is an evolutionarily novel mental apparatus that functions to build justification narratives that legitimize actions and claims. To put this in everyday terms, the self-consciousness system is the language-based portion of one's mind that is narrating what is happening, why it is happening, and why one is doing what in that context. Although this is the epicenter of adult human phenomenological experience, the Justification Hypothesis posits that such justification narratives are evolutionarily recent, emerging in close conjunction with the evolution of language in general. This is important because it suggests that the self-conscious portion of your mind represents only a part of your cognitive processes, one that

can be at least theoretically separated from other cognitive phenomena. As we will see later, this is exactly what neuropsychologists have found.

The adaptive problem (or lock) specified by the Justification Hypothesis is the problem of social justification, which is the problem of explaining the legitimacy of one's thoughts and actions to others. Later I explain in more detail that when the evolution of language reached sufficient complexity, it afforded a window into other's minds, which was a problem because it was not always in one's best interest to allow one's thoughts to be "seen." I further explain why the ability to effectively justify one's claims and actions must have been closely related to the amount of social influence one achieved and was thus tied to reproductive success.

The reverse engineering matching process leads to several predictions. For example, this formulation clearly predicts that the self-consciousness system should be designed in such a way that it allows humans to effectively justify their actions to others in a manner that, all things being equal, tends to maximize social influence. An examination of some of the characteristics of human self-consciousness as elucidated by neuropsychology, social, cognitive, and developmental psychology will be reviewed to demonstrate that there is a large body of general human psychological research that is consistent with this proposition. It will be concluded that the problem of social justification is a prime suspect for a selection pressure that resulted in the evolution of the human self-consciousness system. Said differently, the argument will have been made that the human self-consciousness system can be thought of as the mental organ of social justification.

Before proceeding, it is important to state here that although the Justification Hypothesis is anchored in evolutionary theory, I am of course not arguing that people are born with a fully functioning justification system. That would obviously not be a tenable position. What I am arguing is that humans have a mental apparatus that allows for the growth and development of justification systems with experience. This argument will become clearer as we review lines of research in social, cognitive, and developmental psychology. One way to think about it is to consider that just as evolution has prepared the behavioral investment system with capacities to experience pleasure and pain to guide the manner in which behavioral patterns develop in development, justification systems likewise develop in response to the context of social justification in which the individual lives.

## **Review of the Evidence for Advanced Capacity for Self-Consciousness in Humans**

A key element of the Justification Hypothesis is the idea that humans possess a capacity for self-consciousness that is fundamentally different than that of other animals. This is not a novel proposal. Indeed, many theorists and philosophers have suggested that it is the presence of advanced capacities for self-reflective awareness that differentiates human consciousness from the consciousness of other animals. Daniel Dennett (1996), for example, made the distinction between first-order and second-order intentional beings. A first-order intentional being has a

mental life, consisting of beliefs and desires about many things, but not beliefs and desires about beliefs and desires. In short, first-order intentional creatures are aware, but not aware that they are aware. Second-order intentional beings—namely humans—have beliefs and desires about beliefs and desires; they are aware of their awareness. Likewise, Edelman (1992) distinguished between primary consciousness and higher-order consciousness. He defined primary consciousness as the state of being mentally aware of things in the world, of having mental images of the present. Higher-order consciousness is awareness of the self or the process of being conscious of being conscious. He argued it is intimately tied to language and is only possessed by humans. In one form or another many others have made similar proposals.

In addition to theoretical and philosophical arguments, there have been some empirical investigations of animal self-awareness. Of course, without language such investigations are obviously difficult (a point worth noting in and of itself). However, the ingenious technique of the Mirror Self-Recognition (MSR) task developed by Gallup (1970) has yielded interesting results. MSR involves exposing animals to a mirror, anesthetizing them, and placing a large dot on their forehead to see if they attend to the dot when given the chance to look in the mirror upon awakening. Success at the MSR suggests at least a rudimentary cognitive capacity to become the focus of one's own attention. Success does not mean that the animal can introspect or be self-reflective or has an elaborate self-consciousness system (Mitchell, 1994). On the other hand, failure to succeed at the MSR task is a result that is much easier to interpret and is good evidence for the lack of any genuine capacity for self-consciousness. What is remarkable then is that virtually all animals fail the MSR task. Only adult chimpanzees, bonobos, orangutans, and bottlenose dolphins seem to regularly pass the MSR task. Even most adult gorillas fail to pass the test (Parker, Mitchell, & Boccia, 1994). Thus, the vast majority of mammals do not possess even the most basic cognitive capacities required for self-reflective awareness. Humans generally pass the MSR task between the ages of 18 and 24 months. When one considers how adult humans explain their actions to others, worry about their death, develop grand narratives to account for their existence, and plan their actions weeks, months, and even years in advance, the gulf in self-awareness between humans and even our nearest animal relatives is truly astounding.

Other empirical evidence comes from studies of individuals with brain injury. Neuropsychological investigations provide convincing evidence that human self-awareness requires specific types of information processing systems in the brain that allow for self-conscious awareness. Such investigations have demonstrated that there are linguistically based declarative memory systems in the brain that store information in a manner that allows the individual to consciously remember what happened. Brain damage can also result in gross disturbances in self-awareness, a condition known as anosognosia. Such individuals will often exhibit a remarkable, almost unbelievable, lack of self-awareness, often despite intact intellectual functioning as measured by intelligence tests. Examples include individuals who completely deny that half of their body is paralyzed despite being confined to a wheelchair and unable to move their arm or leg (Barr, 1998). Some cases have

been reported in which the individual recognizes that his arm is paralyzed, but denies that his leg is! Many neuropsychologists explain these phenomena as disturbances in the self-awareness system (e.g., Amador & David, 1998). As explored in more detail below, the language structures of the left hemisphere are thought to be intimately related to human capacity for self-awareness. Anosognosia is often the result of right hemisphere damage, and it has long been proposed that such damage results in the isolation of the language structures in the left hemisphere that allow for self-reflection and self-reporting (Geschwind, 1965).

The enormous difference between human and animal self-awareness and the presence of neuro-information processing systems that allow for self-awareness, suggests that the capacity to be reflectively conscious of our perceptions, thoughts, and feelings is part of the complex functional design of the human brain and mental architecture. A reasonable inference from these observations is that the human self-consciousness system has been shaped by natural selection. However, the case is currently far from conclusive. It is still possible that capacity for self-consciousness is simply a byproduct of other evolved capacities, or is a purely acquired ability, like the capacity to play chess. From a reverse engineering perspective, if a particular adaptive problem could be identified that has been present only in the hominid line and would require an elaborate self-awareness system, then the case that the SCS was the product of natural selection would be significantly strengthened.

## **The Role of Language in the Emergence of the Problem of Social Justification**

Most evolutionists and psycholinguists agree that the human capacity for language evolved through the process of natural selection. These theorists note facts such as humans everywhere possess language; there is a developmental period in which children acquire language easily and rapidly; children learn to speak with remarkably little direct instruction; there are well-documented language processing centers in the brain; and, the vocal chords of humans are elaborately constructed to allow for extremely complex sounds to be generated (Lieberman, 1998). Further, other animals (e.g., chimpanzees) can obtain only a crude approximation of human language despite immense training (Deacon, 1997). Finally, groups of children raised in the absence of a native language have been known to develop a fully functioning language in as little as a single generation (Bickerton, 1995).

In addition to these elements, the capacity for language results in many advantages. It allows valuable information to be shared cheaply and effectively, which in turn allows for more synergistic and cooperative relationships. Language also allows for the accumulation of information across the generations. Furthermore, the ability to symbolically represent perceptual objects and their transformations in the forms of nouns and verbs results in an enormously increased capacity to manipulate, elaborate, connect, and remember concepts. Finally, the absence of useful alternative explanations makes the notion that the human capacity for language is a product of natural selection quite solid.

Crucially for our current considerations, as language sophistication advanced beyond simple descriptions and commands, it came to provide a means to more directly access and assess the thoughts and intentions of others. Although a chimpanzee can clearly send the message that she is angry or scared, without a symbolic language it is almost impossible for her to communicate the reasons why she feels that way. Humans are different. Unlike chimps, language allows humans to ask and be asked about the thought processes associated with their behaviors. Questions such as “Why did you do that?; What gives you the right to behave that way?; Why should I trust you?” force the issue. Obtaining information about what others think, what they have done, what they plan to do is obviously important for navigating the social environment in modern times and given that humans have always been an intensely social species, there is every reason to believe that it was equally essential in the ancestral past. So once people developed the capacity to use language to access the thought patterns of others, they likely did so with vigor.

Now consider why the answers to those questions would have been so important. If you strike a comrade with a stick, it matters whether you tell him it was done by accident or on purpose. If you take more than your proportional share of meat, it matters how you explain that action. If you are bargaining with a stranger, you can get more resources if you emphasize why the resources you are trading are valuable, and so on. In short, a second assumption associated with the problem of social justification is that the kinds of explanations people offer for their behavior have real world consequences. If you question the validity of this assertion, I encourage you to alter your explanations to others and see what happens. For example, next time you spill your drink on someone, tell her that you meant to do it, and see if she responds differently than if you had told her it was an accident.

A third assumption underlying the problem of social justification is that human interests diverge and this complicates the translation process significantly. If one’s interests always fully coincided with the interests of others, communicating the reasons for one’s behavior would primarily be a technical problem of translating one’s nonverbal thoughts into a symbolic form that could be understood. But because interests always diverge to some extent and the explanations given for one’s behavior have real world consequences, the communication task becomes one of justification rather than simple translation. Elaborating a little bit on this point with an example, consider what would happen if a male (‘Beta’), was interested in forging a relationship with a particular female, but she was paired with another male (‘Alpha’). Imagine further that Beta starts spending time with her but is then confronted by Alpha with a question such as “Why are you spending so much time around her?” If Beta simply translated his thought processes in response to the question, he would say something like: “I want to separate the two of you and have her as my mate.” Of course, such a statement seems blatantly foolish because the information would obviously be of crucial importance to Alpha, who, upon hearing it, would be able to take defensive action.

From the vantage point of the Justification Hypothesis, the problem is how Beta can explain his actions without costing him vital social influence and opportunities. To do this effectively, he must reflect on his actions, take into account the interests

and knowledge of his audience and develop a justification narrative that provides a plausible explanation of the evidence without costing him key resources. In this light a response such as, “She is teaching me how to plant seeds” might be a good justification, in that it could provide an explanation for Beta’s actions in a manner that avoids potentially negative social consequences. Of course, Alpha must then evaluate the coherence and consistency of such an explanation and decide whether or not to accept the justification. Importantly, the need to evaluate others’ justification systems would have created selection pressures for an enormous increase in “theory of mind” abilities, which are the abilities to infer the intentions of others and are one of the most advanced capacities humans demonstrate (Shaffer, 2008).

To summarize the key points so far: (1) Humans have advanced capacities for self-reflective awareness that far outstrip other animals, and these capacities are dependent on particular structures in the brain that are more developed in humans than other animals; (2) As language increased in sophistication, it must have reached a tipping point whereby it allowed others access to one’s thought processes; and (3) This access to one’s thought processes created a problem to the extent that interests between individuals diverge because in such instances it put pressure on individuals to create socially justifiable explanations, not simply share their thought processes as accurately as possible. A fourth and final assumption of the Justification Hypothesis is that, given the above listed assumptions, standard evolutionary processes will operate to give rise to increasingly sophisticated capacities for social justification. Specifically, it is presumed that as better justifiers emerged, an evolutionary arms race would have ensued, resulting in increasingly adept justifiers and increasingly more sophisticated capacities to analyze and critique the justifications of others. Moreover, it will be argued later that justification systems themselves evolve.

## **A Tripartite Model of Human Consciousness**

Up until this point in the argument, the Justification Hypothesis can be considered a good Just-So Story (Gould & Lewontin, 1979). That is, I have generated a plausible “story” regarding human evolutionary history and possible selection pressures that resulted in the human self-consciousness system. Yet the value of the Justification Hypothesis is not simply in that it provides a plausible story for why the human self-consciousness system might have evolved. Instead, the real value is found in the light it sheds on understanding human consciousness and culture in the current context. For starters, I have found the Justification Hypothesis helps clarify the nature of human consciousness from a more clinical, phenomenological, and folk psychological perspective. That is, it has given rise to a rather straightforward formulation about the different domains of human consciousness that I find to be qualitatively generalizable. As a clinician and educator, I have used the formulation with much success in both understanding and guiding interventions with my patients and supervising my graduate students. Virtually everyone with whom I have shared the conception can easily relate to it. Below I show how adult human



consciousness can be readily understood as consisting of three domains that emerge from the logic of the Justification Hypothesis: (1) an experiential consciousness system; (2) a private self-consciousness system; and (3) a public self-consciousness system.

Before we proceed, however, we need to once again make sure we are clear on the meaning of our terms and the relationship and differences between cognition, consciousness, and the mind. The previous chapter introduced a schematic of the architecture of the human mind, and the focus was on the way in which information was encoded and processed by the nervous system. In this next section, the focus is on consciousness. Consciousness refers to the first person, phenomenological phenomena of experiencing and awareness. I define consciousness as an “experienced” electro-neuro-chemical representation of animal–environment relationships. Experienced is in quotes because it highlights that the fact that the processes by which neurocognitive events gives rise to felt experience remains one of the great mysteries in psychology and neuroscience (Koch & Greenfield, 2007). An important point about the definition is that it means that consciousness is a specific kind of cognitive process. All conscious phenomena are thus cognitive phenomena, but cognitive processes can be either conscious or nonconscious (or implicit or explicit in the language of some cognitive psychologists). Although often thought of in dichotomous terms, it is also important to note that we can place cognitive processes on a continuum of consciousness, from completely nonconscious to somewhat conscious to fully self-conscious. Researchers have documented much information is processed completely nonconsciously. And when we consider altered states of consciousness (either induced by drugs, diseases, or hypnotic trances), dreaming, day dreaming, and transitioning from sleep to wakefulness, we can see the dimensional aspect of conscious awareness.

A second consideration relating cognitive and conscious processes is seen when we consider the issue of memory and stored information. When we think about all the information that we can remember and contrast it to what we can be conscious of at any particular moment, we can clearly see that consciousness is not synonymous with all the information stored in the mind. Freud called information stored in memory that is potentially accessible the preconscious. The preconscious is material that can be freely brought into conscious but is not currently the focus of attention. So you have stored memories about your last meal or of your parents or an important accomplishment that can be brought into consciousness either via external stimuli or self-initiated efforts to recall.

And then there is the domain of subconscious processes. Subconscious processes are experiences that are at the edge of self-awareness but do not receive enough allocation of attention for them to enter full consciousness. This might be because the stimuli or experiences were not salient enough to warrant the attentional resource or because the individual was motivated to shift attention away from them. We will have more to say about the latter process when we introduce the concept of filtering. But if you have ever had the experience of trying not to think about something or have been guided to experience something you knew was inside you all along but previously could not allow yourself to be conscious of, then you are familiar with

the idea of subconscious processes. With these definitional issues addressed, we can turn to looking at human consciousness from the lens of the unified theory.

## A Dual Processing Model of Cognition and Consciousness

The unified theory strongly supports a two domain view of the human mind, which is the idea that human mental architecture consists of two kinds of information processing systems. The first system, as framed by Behavioral Investment Theory, is a nonverbal, perceptual–motivational–emotional (i.e.,  $P - M \Rightarrow E$ ), parallel neuro-information processing behavioral guidance system that computes resource availability and organizes action. The second system, framed by the Justification Hypothesis, is uniquely human and is a verbal (symbolic–syntactical), reflective, logical–analytic, sequential information processing system. Importantly from the perspective of integrating psychological theories, dual processing models of the human mind are found in work in psychotherapy, neuropsychology, cognitive, and social psychology. In fact, so fundamental is this general conception of two broad mental domains that it has been proposed as the basis for a central dogma in human psychology (Cook, 1989).

Because consciousness is a particular form of cognitive process and because humans have two distinct forms of information processing systems, it follows that humans have two different forms of consciousness. And, indeed, there are two broad categories of mental experience—domains that can be referred to as sentience and self-consciousness. Sentience is the term used to describe nonverbal conscious experiences, such as feeling pain, seeing red, being hungry, or imagining an event. In contrast, self-consciousness refers to the language-based, self-reflective thought. Whereas the former involves experiencing (e.g., feeling hungry), the latter involves self-awareness and verbal making-meaning about that experiencing (“Here I am, feeling hungry, and wishing that I could get home for dinner”).

Epstein (1994) proposed a two domain model of the human mind and consciousness that integrates psychodynamic theory with cognitive science in a way that is very congruent with the unified theory. Epstein labeled the first domain the “experiential” system. The experiential system processes information relatively automatically, holistically, in parallel and via images. It is self-evidently valid, as in “seeing is believing.” He labeled the second domain the rational system, which processes information sequentially, via words and symbols, is reason-oriented and requires justification via logic and evidence. Epstein reviewed a large body of evidence congruent with this model of consciousness and pointed out how the model is consistent with folk psychological notions of the distinction between thinking and feeling, head and heart, and reasoning and intuition. It is also congruent with observations about people with irrational phobias. Consider how these individuals “know” at a rational level their fears are unjustified but are nevertheless unable to act differently because of the power of the experiential system.

A gap in Epstein’s innovative model—filled in by the Justification Hypothesis—is a conceptual frame explaining *why* the so-called rational mind evolved in humans

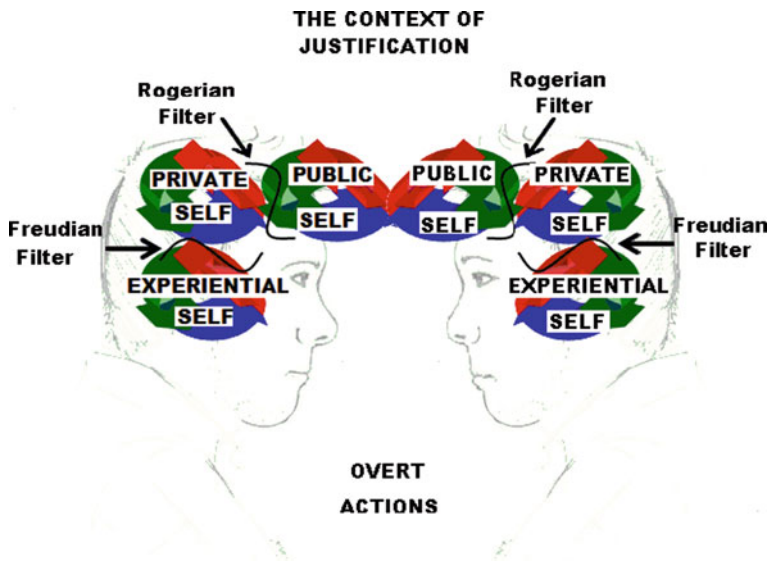
and *why* it exhibits the design features it does. Epstein explicitly stated his model was congruent with a modern psychodynamic perspective, and the Justification Hypothesis enriches this picture significantly because it characterizes the second domain as a justification system that inhibits and filters out nonverbal thoughts, images, and impulses that are painful or unacceptable and allows justifiable actions to be expressed. Another related area that results in extending Epstein's model is because the Justification Hypothesis emphasizes the important role of social justification, it leads to insights about the difference between the private and public self-consciousness system.

To see clearly how the Justification Hypothesis does this, we can turn to another influential model of the self-system constructed by a psychologist interested in ego development. Loevinger (1976) defined the ego as the self-system that (a) manages impulses; (b) allows individuals to understand and explain themselves to others; and (c) allows individuals to experience themselves as a coherent entity over time. She argued that as children grow into adolescents and adults, their egos develop into increasingly complex networks of justifications regarding themselves, others, and the world around them. An example can help illustrate some of these elements. Several years ago I was discreetly observing my 4-year-old daughter, Sydney, playing with my 2-year-old son, Jon. I watched as the playing escalated into a more direct competition, which resulted in Sydney pushing Jon down, resulting in a cry of distress. Appearing on the scene and seeing Jon crying, I somewhat angrily asked Sydney what had happened. Almost immediately, she replied "he fell." As this example illustrates, impulses are not always controlled and incriminating information (i.e., the fact that she pushed him) is often filtered out quite automatically.

From the vantage point of the Justification Hypothesis, the first two aspects of Loevinger's model (i.e., impulse control and explaining one's self to others) highlight two distinct steps in the justification process. The first step involves the process of translating the nonverbal images into the self-consciousness system and blocking those impulses that are problematic, painful, or not consistent with the individual's identity. The second step involves determining whether and how to share one's private thoughts with others. This two-step process highlights an important component implicit in the Justification Hypothesis; namely that for adults, there are two domains of justification: the private (our internal self-talk) and the public (what we say to others). With this separation highlighted, let me introduce a figure that maps out the essential components of human consciousness, social context, and overt behavioral investments.

## The Three Domains of Human Consciousness

As depicted in Fig. 5.1, there are three broad domains of human consciousness: (1) the Experiential Self; (2) the Private Self-Consciousness System (the Private Self); and (3) the Public Self. The experiential self refers to the sentient aspects of consciousness, and it is made up of the qualia or the "raw feels" of conscious experience. Consistent with the  $P - M \Rightarrow E$  formulation, these experiences can



**Fig. 5.1** The three domains of human consciousness

be generally classified into sensations and perceptions (e.g., seeing red, touching a rock), motivational urges (e.g., hunger, sexual desire), and feelings and emotions (e.g., sadness, joy, anger), as well as imagined objects or occurrences. The two other domains of human consciousness represent the two separable domains of justification, the private and the public.

The private self is the center of self-reflective awareness in adults, and it is made up most immediately of the internal dialogue that weaves a narrative of what is happening and why. It is a second-order awareness system, one that translates and feeds back onto the experiential system. Psychodynamic theorists generally consider this the conscious portion of the human ego. It also is the part of the mind targeted by traditional cognitive psychotherapy, which teaches individuals to monitor the content of their private justifications, identify how those justifications influence feeling states and behavior, and develop strategies for analyzing the accuracy and utility of those justifications in promoting adaptive action (Henriques, 2007). The public self exists between individuals and is the explicit articulation to others of what one thinks, along with the image one tries to project. It is my contention that these three domains of consciousness should be readily identifiable by every healthy human adult.

Early in the developmental sequence, the private and public justification systems are not clearly separated. As language develops, specific actions are either inhibited or allowed depending on the strength of the rule, the magnitude of the impulse, and the development of executive functioning. Although this is indeed a justifying filtration process as illustrated by the anecdote of my daughter described earlier, it is not genuinely reflective in early childhood (i.e., ages 2–3), as it does not involve the

clear separation of one's own perspective from that of others. The private justification system emerges as dialogue is internalized, and in later childhood and certainly by early adolescence there is a distinct psychosocial identity that becomes the seat of reflective self-awareness in adults. The following quote from Carl Jung captures this emergence vividly:

I was taking the long road to school. . .when suddenly for a single moment I had the overwhelming impression of having just emerged from a dense cloud. I knew all at once: I am myself! . . .Previously I had existed, too, but everything merely happened to me. . .Previously I had been willed to do this and that: now I willed. This experience seemed to me tremendously important and new: there was "authority" in me. (Ryckman, 2004, p. 75)

The public self is a mixture of how we want to be seen and how we imagine we are seen by others, although both may be quite different from how one's presentation is actually received by others. A number of seminal theorists have emphasized the importance of and dynamic tension between the public and private identity. James (1890, p. 294), for example, argued that identity was so intimately tied to the social world that people have "as many different social selves as there are distinct groups of persons about whose opinion he cares." Cooley's formulation of the "looking-glass self" and Jung's concept of the persona are other well known examples of early theorists emphasizing the role of others in shaping our identity. From the perspective of the Justification Hypothesis, one of the central tasks of navigating the social environment is maintaining a justifiable position in the eyes of others.

The microsociology of Erving Goffman (1959) makes a strong case in favor of the importance of the public persona in a way that is very consistent with the Justification Hypothesis. In *The Presentation of Self in Everyday Life*, Goffman describes face-to-face interactions and examined such processes through the lens of stage acting. He articulated how interpersonal interactions could be considered "performances" as actors learned to *manage* their impressions to others in both the structured and improvised roles of everyday life. Specifically, Goffman suggested that actors work to convey a positive, predictable impression, so as to be perceived as justifiable in the eyes of the audience. Interestingly, Goffman argued that in cooperative settings, actors work to preserve the justifiability of one another, and frequently move away from trajectories that could be unflattering or embarrassing for either party in the interaction. Over the years, his writings and the work of others provide a rich catalog of strategies and tactics that performers employ with the hope of managing the impressions others will form of them.

The relationship between the public self that people attempt to project and how it is received is a crucial element of interpersonal relations and mental health. As you undoubtedly can imagine and likely have experienced, disconnect between the image we attempt to convey and the image received can happen for a multitude of reasons. We can misread what others want or expect, we can simply fail to impress despite our best efforts, and we often need to manage conflicting interests, both ours and those with whom we relate, and this can become especially complicated when we are aligned to interpersonal systems that are in conflict. Consider, for example, when one friend complains to you about another. In supporting one

friend, tension emerges with the other. Moreover, lying, impression management, saying one thing and doing another are common everyday occurrences, so other people are naturally attuned to the possibility of filtering, deceit, and incongruence between actions and words, and are constantly wondering about what people are really thinking compared to what they are socially presenting. These issues make the social world complicated, which in turn results in a heightened self-focused attention for many. Researchers have documented empirically the distinction between the two domains of self-focused attention (the public and the private), have shown that individuals differ greatly in the extent to which they focus on one or the other or both, and have demonstrated that high levels of self-focused attention in one or both domains is associated with anxiety, depression, and other forms of psychopathology (Fenigstein, 2009).

## The Context of Justification

The three domains of consciousness are not the only aspects in Fig. 5.1. Above the two figures is labeled “The Context of Justification,” which refers to the network of symbolically based beliefs and values that provide the interacting members a shared frame of reference for their interaction (cf., Reichenbach, 1938). The context can be considered on the dimensions of time and scope. Scope refers to the size and scale of the context one is considering. Uri Bronfenbrenner’s (1979) Ecological Systems Theory provides a useful framework for considering the scope of the system, although I should note that he was concerned with the whole societal context, which would include the biophysical ecology and technology, in addition to the systems of justification. Nevertheless, his levels are useful for framing our perspective on the context of justification, and thus I briefly review them here.

Bronfenbrenner delineated four separable, but interrelated and nested contextual levels: (1) The *microsystem* consists of the immediate relational environment that individuals interact in, such as the home or school; (2) The *mesosystem* refers to the interconnections between the various microsystems (such as the interaction between school and home), and this level can also be thought of as the general community in which the individual lives; (3) The *exosystem* refers to the larger, distal contexts that the individual does not necessarily directly experience in, but can nevertheless have an influence (for instance, the economy shifts and a child’s parent loses her job); and, (4) the *macrosystem* refers to the larger cultural context, especially the beliefs, values, laws and mores that define the culture as a whole.

Time is the other dimension on which the context of justification needs be analyzed, as all justification systems arise out of a historical context and are legitimized in part by that history. Bronfenbrenner also integrated this dimension into his model, labeling it the fifth “system,” which he called the *chronosystem*. By considering the time dimension, we can place the immediate justifications in a historical and developmental context. Thus, if we are considering someone offering an excuse for a dishonest act, the history of making such excuses will play a large role in how the justification is received by others.

In addition to applying to the more immediate context of individuals giving reasons to one another as in the clip from *Ordinary People* highlighted in the previous chapter, the Justification Hypothesis also readily applies to the macro-level cultural context, which consists of the laws, religions, and political positions that exist at national levels and beyond. The religious, legal, and normative systems of social convention, situated as they are in historical trajectories, all provide the larger context in which the specific actions and scripts of local individuals are played out. Because all these systems broadly frame which actions are legitimate and which are not, they can readily be considered as large-scale or collective justification systems. For example, consider racial attitudes in America through this lens. Prior to the 1950s in America, there was a general legitimization of the racial superiority of Whites by the people in power, such that in many circumstances it was acceptable to publically denigrate minorities. However, with the rise and relative success of the Civil Rights Movement, explicit racism became generally unjustifiable and now such pronouncements, if made in public are appropriately met with scorn and punishment. A host of other possible examples are available and later in the chapter I articulate how religion can be readily conceived as a large-scale justification system.

Actions are also labeled in the figure, and are fairly straight forward. These are the set of observable behaviors that the individuals engage in, and are explicitly defined as the functional changes between the individual and the environment (as opposed to changes within the individual). Actions thus are overt mental behaviors. In more colloquial terms, we can think of actions as the practices people engage in, or the work of everyday living. Thus, if you are cooking a meal, taking a trip to the emergency room, paying the bills, or making love, these are the practices or actions you are engaged in. If we consider the major variables highlighted in the picture, we can see that people must navigate and attempt to align their justifications with their practices in particular relationship and social contexts.

## The Two Filters

Inside each of the individuals in the figure are two filters, labeled as the Freudian and the Rogerian. The Freudian filter exists between the experiential self and the private self and refers to the process by which unjustifiable or painful images and impulses are filtered out and/or are reinterpreted to be consistent with the individual's conscious justification system. It is called the "Freudian" filter because the dynamic relationship between self-conscious and subconscious thoughts and feelings was (and still is) a central focus in both classical psychoanalysis and modern psychodynamic theory. Indeed, I have argued that the idea that the self-consciousness system filters out socially unacceptable motives in a manner that allows for a defensible conception of the self is arguably the founding insight of Freudian psychology (Henriques, 2003). Consider, for example, that repression and rationalization, two central defense mechanisms, can be readily thought of as two sides of a filtering process. Repression blocks or filters out unacceptable experiences; and rationalization

is the process of attempting to provide an acceptable verbal narrative for those actions that are expressed.

Why are certain impulses filtered? According to the Justification Hypothesis, the reason is to maintain a consistent, relatively stable justification narrative of the self and to maintain a justifiable image in the eyes of others. In his book *Ego Defenses and the Legitimization of Behavior*, Swanson (1988) made exactly this point, explicitly arguing that we should think of all ego defenses as “justifications that people make to themselves and others—justifications so designed that the defender, not just other people, can accept them” (p. 159). The formulation remains a hallmark feature of modern psychodynamic models and is also present in general models of personality and social psychological research on cognitive dissonance and psychopathology. The current understanding of this filtering process from a modern psychodynamic view is nicely captured by the Malan Triangle of Conflict shown in Fig. 5.2 that depicts impulses/feelings at one point, anxiety at another, and defenses at the third.

The Malan triangle represents how impulses and feelings begin to emerge, but those that are painful, problematic, or socially unjustifiable result in signal anxiety, which in turn activates a defense mechanism like repression or rationalization to avoid the threat and restore psychic equilibrium. Psychodynamic conceptions of anxiety and defense can thus be thought of as part of the filtration processes that regulates the relationship between affectively laden subconscious experience and self-conscious thought. Seen in this light, the classic psychoanalytic technique of free association, whereby the primary instruction for the analysand is to say whatever comes into his or her mind, fits well with this formulation in that it speaks to how Freudian theorists conceptualized the process as a filter that required a special context and instruction for it to be turned down, resulting in access to the blocked material.

To help clarify how the relationship between unjustifiable images and thoughts, affectively laden experiences, anxiety, and defenses consider the following imaginary exchange between husband and wife:

Dan is a 43-year-old high school football coach who has a masculine identity and beliefs about the importance of self-reliance, mental toughness, and self-assurance. Despite this conscious identity, it is also the case that Dan’s mother died when he was 8, and he struggled with unmet dependency needs during his early years. Fifteen years ago, Dan married Janice, and they had a fairly traditional marriage, where Dan worked and Janice raised their two kids who are now 12 and 10. However, six months ago, Janice began a new job in marketing,

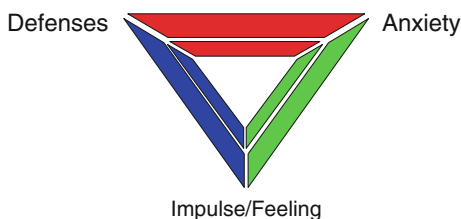


Fig. 5.2 The Malan triangle of conflict



which Dan initially supported. She is currently traveling to make her first major marketing pitch and is away for four days. She calls him after the first day to share her excitement that her portion of the presentation went well.

Janice excited, "Hi, Dan. You are not going to believe this, but I did it and did it well! I am serious, I really impressed them."

Dan, in a monotone voice, "That is great, honey. What time did you say you are coming home Thursday?"

"Around 5:00 pm. Anyway, I gave them the idea about nature and connecting the materials to the green revolution, and they thought that it was an excellent idea that I should definitely develop further. Isn't that great? Oh that reminds me, could you water the plants I just seeded? I am worried they will die if they don't get water every other day."

"If I think about it," Dan replies vaguely.

"What is that supposed to mean?" Janice says, her tone of voice changing to anger. "It is important. Actually, my presentation was important! What is wrong with you?"

"Nothing at all." Dan responds, in a somewhat irritated tone.

"I don't know what it is, Dan, but I swear you undermine me sometimes."

"Whatever. . ." Dan says, dismissively.

"Fine, and thanks for all your support." Janice says sarcastically and hangs up.

Dan gets off the phone, thinking that Janice was in a lousy mood.

With this vignette, we can readily imagine the dynamics that are taking place in Dan's intrapsychic system. On the one hand, Dan has an image of himself as self-reliant and self-assured and thus when Janice started to go back to work, consistent with his own identity, he said it was fine. However, her redirecting her time and energies elsewhere activated in him some subconscious feelings connected to the unmet dependency needs prominent in his childhood. Yet, these memories were both painful and challenged his own private justification system for who he was and how he should be. Consequently, as these images and affects began to take form in his consciousness they were associated with signal anxiety, which led him to avoid and repress (defense mechanisms) them.

The conflict begins to boil on the phone because Janice is reporting success in her job, which Dan subconsciously knows will mean more time away. And yet, given that she is his wife, that he supported her decision to go back to work, and that he sees himself as a self-assured and self-reliant man, he consciously feels he should be happy for her, and the anxieties that her success activates in him cannot be explicitly justified. And yet his underlying feelings of unmet dependency move him to want to avoid supporting her success. Thus his attention shifts, initially landing on when she is coming home. When she continues and asks him to do something for her, he is vague and noncommittal, although he doesn't say he won't do it because that would not be justifiable. Sensing something amiss, she asks him if something is wrong. Dan, of course, cannot even tell himself what is wrong, and so certainly cannot publicly share his thoughts with Janice. Consequently, he engages in denial and implicitly suggests Janice is seeing things that aren't there. Janice attempts to call him on his lack of support but because he has been vague, he can deny that accusation and attribute the bad turn the conversation took to Janice being in a bad mood, projecting the negativity on to her and explaining away the conflict.

While the focus in the example was on the filtering between the experiential and private self-consciousness system, there is, of course, also filtering between the

private and the public. The film *Liar* starring Jim Carey provides an excellent illustration of the nature of the private to public filtering. In that film, Carey plays an unscrupulous but effective lawyer who is always making promises to his son, but fails to follow through. After failing to show up for his son's birthday party, his son makes a wish that his father cannot tell a lie for 24 hours that magically comes true. The film is a situation comedy that shows what would happen if we could not filter our private thoughts, but instead when asked, we had to share our private justifications in an unabridged manner with others. I call the filtering that takes place between the private and the public self the Rogerian filter because in relationship to early psychoanalytic thinking, Rogers shifted the focus from deep and largely subconscious intrapsychic processes to more conscious thought and experiences and here-and-now interpersonal processes. He emphasized that the root of much psychopathology was in how judgmental others would stunt the development of one's "true self." This is because, fearing judgment, individuals filter out their true desires and put on a mask—a "social self"—to appease influential others. Person-centered therapy is based on the premise that through forming a relationship with an empathetic, nonjudgmental other, individuals can stop the problematic stunting caused by the private to public filtering process, reinvest in their true sense of self, and return to a path of growth and fulfillment. This said, I do want to note here though that Rogers was also very concerned with filtering between the private self and experiential consciousness and many modern day neo-Rogerian therapies like Greenberg's (2002) Emotion Focused Therapy focus much of their attention on this aspect of intrapsychic functioning.

Social psychologists have also shown much interest in the filtering between the public and private selves. For example, Snyder (1974) explored individual differences in amount of filtration people would engage in. He termed those who tended to filter their private thoughts and work harder to share the beliefs and values of the people around them as "high self monitors." Those individuals who were more likely to "speak their minds" and focus more on their own opinions he characterized as "low self monitors." Numerous empirical investigations have confirmed the importance of this variable in influencing behavior in various social circumstances. Jackson (1988, p. 121) offered the following characterization of what the research in social psychology has found:

People assume that if their behavior can be observed that it will be interpreted and evaluated. They attempt to avoid negative evaluation in the process of constructing their conduct, before it "emerges" for public interpretation. In a social situation, participants monitor others' reactions to their own conduct. Their behavior acquires a situated meaning, specific to the social act. If one's conduct means something different to others than what one intended or assumed, one makes verbal and behavioral adjustments until one obtains the desired reactions.

An example of research into repression serves well to highlight both the Freudian and Rogerian filtering processes and the dynamic interplay among the experiential self, the private self, and the public self. To explore the dynamics of sexual guilt, Morokoff (1985) divided a sample of women into "high sexual guilt" and

“low sexual guilt” groups and had them watch an erotic film for 12 minutes. Each woman’s arousal was measured physiologically, and they were also given self-reports. The results indicated that the high-guilt women experienced more physiological arousal but reported less arousal than the low-guilt women, and this was interpreted as evidence of repression. Specifically, the data from the physiological measure was offered as evidence that a subconscious portion of their experiential minds responded to the explicit material with relatively high levels of sexual arousal, and yet they ostensibly did not have self-conscious access to such arousal, presumably because their identity judged such arousal to be unjustifiable. If this was the case, this would be the Freudian filter at work. Of course, it could be that the women consciously recognized their arousal but did not want to publicly share that with the experimenter and intentionally minimized their self-reported arousal due to social demand characteristics. If so, this would be an example of the Rogerian filter at work.

In summary, the Justification Hypothesis, in conjunction with Behavioral Investment Theory and the Influence Matrix, leads to a new tripartite model of human consciousness that consists of an experiential self, a private self, and a public self, along with two distinguishable filtering processes. The model is derived first from the notion that the experiential consciousness evolved as part of the behavioral guidance system, whereby pleasure and pain serve as signals to approach and avoid, and images serve to simulate behavior so that the animal can anticipate outcomes without expending the energy on the overt behavior. Then symbolic language, a new system of information processing, emerged. Understanding the language system as being functionally organized as a justification system that evolved in response to the problem of social justification, we deduced quite clearly why there would be two domains of justification—the public and the private—and that there would be filtering processes among the three domains of consciousness. Crucially, exactly the kinds of filtering processes suggested by the logic of the Justification Hypothesis are found in psychodynamic theory, Rogerian theory and therapy, cognitive science, and social psychology.

## **Assimilating and Integrating Lines of Research with the Justification Hypothesis**

The Justification Hypothesis was the first piece of the unified theory to be developed, and it was because I saw processes of justification everywhere I looked, combined with the fact that the idea seemed to be able to assimilate and integrate disparate lines of research in cognitive, developmental, and social psychology that started me down the path of developing a new way to unify psychology. In the subsequent section, lines of research on the interpreter function of the left hemisphere, cognitive dissonance, self-serving biases, the organizing forces in human self-knowledge, implicit and explicit attitudes, defense mechanisms, how people give of excuses and accounts of their behavior, and reason giving processes are reviewed that together show why the Justification Hypothesis is an integrative hub of an idea.

## The Justification Hypothesis and the Interpreter Function of the Left Hemisphere

The Justification Hypothesis posits that the capacity to justify is a unique and relatively recent addition to hominid mental architecture. An implication of this assertion is that we should be able to identify the portion of the brain responsible for reason giving and that such processes should be separable from other mental processes. The reason for this claim is that evolution builds on existing structures and because the capacity to justify evolved rather late in the game, we should be able to identify it as such. A unique circumstance in the field of neuropsychology actually allowed this implication to be explicitly examined, and there is now clear evidence of a separable, interpreter system located in the left cortex.

Some individuals experience severe seizures, which result from excessive, uninhibited neural firing that spreads throughout the brain. Beginning in the 1950s and 1960s, neurosurgeons began to cut the corpus callosum in the brains of some patients with severe seizures in an attempt to minimize the spreading of the out-of-control neural firing. The corpus callosum is the set of neural fibers that connects the two hemispheres of the brain, and thus when it is cut, communication between the two hemispheres is broken. These patients—who came to be called split-brains—generally lived normal lives but careful research revealed some striking findings. For example, some patients would report a condition called “alien hand syndrome,” in which the left hand (guided by the right hemisphere) would seemingly act as if it were controlled by a mind of its own (Scepkowski & Cronin-Golomb, 2003). A patient with alien hand might go into the closet to get a blouse, and find that both her hands would reach for separate garments, and there might literally be a tragic-comic struggle between them!

Roger Sperry began to systematically research the consequences of split-brains. He devised a technique for sharing information with only one half of the brain, usually by presenting it to only the right or left visual field. These studies added findings to the idea that the brain is quite lateralized, meaning that the two halves of the brain specialize in different functions, with the left hemisphere being more linguistic and rational and the right being more spatial and visual. Sperry’s student Michael Gazzaniga (1985) began to systematically study how split-brain patients would explain their actions. He reasoned that since it houses the language center, the left hemisphere would be the seat of self-explanation. He wondered what would happen if information was given only to the right hemisphere and people then acted on it. How would people explain their actions in the absence of the correct information? The Justification Hypothesis argues that the human self-consciousness system functions first and foremost as a social justification system, so the prediction is that people should confabulate socially acceptable justifications even when the system does not have access to the necessary information. This is exactly what happened. Gazzaniga (1992) found that if simple commands were flashed to the right hemisphere, such as “walk around” or “laugh,” the patients would follow these commands (the right hemisphere does have rudimentary linguistic capacities). However, when asked to explain why they were performing these behaviors (e.g., walking or

laughing), patients would confabulate a reason, and say “I am going to get a drink” or “Because you guys are so funny.” In other words, their self-consciousness system justified the behavior in the absence of necessary information.

Gazzaniga characterized the system of cognitive processes that allows for these interpretations to occur as “the Interpreter.” He wrote

It is easy to imagine selection pressures promoting an interpreter mechanism in the human brain. A system that allows for thought about the implications of actions, generated by both others as well as the self, will grasp the social context and its meaning for personal survival. . . .Also, the interpreter function generates the possibility for human uniqueness. . . . I think that the built-in capacity of the interpreter gives each of us our local and personal color (Gazzaniga, 1992, p. 134).

That the self-consciousness system appears to be designed in such a way that it develops socially justifiable interpretations of behavior is an important piece of evidence consistent with the Justification Hypothesis.

## The Justification Hypothesis and Cognitive Dissonance

Cognitive dissonance is a classic area of research in social psychology. Initiated by Leon Festinger, one of the most influential social psychologists in history, research into cognitive dissonance involves examining how people experience tension when they hold psychologically inconsistent beliefs and the things people do to resolve that tension. An example of two psychologically inconsistent beliefs would be to believe that one is a good, helpful person and then observe one’s self acting in an unhelpful manner. According to cognitive dissonance, individuals will experience tension if they were to hold such beliefs simultaneously and that they will then work to change their beliefs to reduce the tension.

To get a flavor for research on cognitive dissonance imagine the following situation. After completing an extremely boring task for a psychology experiment, the experimenter asks you to do her a favor. She tells you that she is examining how expectancy biases influence people’s experiences. The next participant for the study has arrived. Unfortunately, her graduate assistant—who was supposed to inform the participant that the task is exciting and enjoyable—is not there and she needs someone to fill in. She then offers you either \$1 or \$20 to tell the student the task was great. After you comply, she then asks you what you really felt about the task. As is now well known, if you received \$1 you rate the task as more enjoyable and less boring than if you received \$20. Why, according to cognitive dissonance, do people do this? Cognitive dissonance researchers frame this in terms of insignificant justification. Telling a fib for \$20 makes sense. However, telling a fib for a small reward does not; there is not sufficient justification for the act. So, individuals reduce the tension by altering their beliefs.

The lens of the Justification Hypothesis and its focus on the problem of social justification adds that what people are doing when they reduce dissonance is working to maintain a justification narrative that allows them to frame their actions in an acceptable manner. This can be seen if one extends the scenario. For example,

imagine that 2 hours after the experiment you are walking around campus and you come across the participant you lied to. “Hey,” he calls, “that task was boring as anything, why did you lie to me and tell me it was exciting?” If the experimenter gave you \$20, you have a reasonable justification and might respond, “Sorry, but it was something they were going to tell you anyway, and she gave me \$20 to say that.” If however, you only received a \$1, it is much harder to justify that you lied. The excuse “she gave me a dollar to lie” is not valid. Yet, if you did not find the task to be so bad, you could defend yourself as follows: “The experimenter asked me to say that. Besides, I didn’t think the task was so bad.”

What cognitive dissonance researchers have discovered is that when confronted with such evidence about the self, people alter their belief systems in a manner that better justifies their previous action. Decades of study and hundreds of experiments have demonstrated that such a process is hugely influential in how people form beliefs about themselves. It is now very clear from social psychological research that people are motivated to justify their actions, beliefs and feelings, and they are acutely attuned to the possibility of being in an unjustifiable state. Moreover, modern researchers have honed in on what exactly is meant by Festinger’s concept of psychologically inconsistent beliefs and now emphasize the concept of *self-justification* (e.g., Aronson, 2007). Directly consistent with the Justification Hypothesis, self-justification refers to the process by which people are motivated to construct justifiable narratives of their actions, beliefs, and feelings.

The dramatic and pervasive tendency of people to engage in self-justification and the powerful impact such processes have on everything from politicians who will not apologize for mistakes to torturers who feel no guilt to therapists who are certain they are uncovering repressed memories to physicians who are unduly influenced by drug companies is wonderfully illustrated in Tarvis and Aronson’s (2007) book, *Mistakes Were Made (But Not by Me): Why We Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts*. From the Milgram experiments to the Watergate scandal to vicious cycles in marriages that lead to ugly divorces, the authors show how time and again the pressures of self- and social justification coupled with blind spots resulting from dissonance reduction can lead individuals down a slippery and dangerous slope, where they find themselves engaging in acts or supporting beliefs they never would have done initially.

Importantly, the processes that function to maintain self-justification and reduce cognitive dissonance generally operate outside of self-awareness. For example, if we were to ask individuals in the above-mentioned experiment what went through their minds, few—if any—would respond, “I initially felt that the task was boring, but then when I found myself willing to lie about it for only a dollar, I realized that this made me vulnerable to attack and criticism for committing a fairly unjustifiable act with insufficient reward. As such, I changed my belief in how boring the task was so that I would be in a better place to justify my actions.” People are conscious only of the result of the dissonance reduction process, rather than the process itself. Likewise, children are not taught directly about dissonance or about how to adjust their beliefs accordingly. Indeed, the whole process is implicit, and it took social psychologists to document its occurrence. The findings associated with

cognitive dissonance research are obviously directly consistent with the Justification Hypothesis, and the implicit nature of cognitive dissonance strongly suggests that it is a design feature of the human mind.

Also important from the vantage point of the Justification Hypothesis is that there is currently no good foundational theory for *why* people engage in dissonance reduction. For example, leading dissonance researcher Joel Cooper (2007, p. 87) recently commented:

Where does dissonance come from? Why do we suffer an unpleasant tension state when, for example, we convince someone to believe in an issue we do not believe, suffer to achieve a goal, choose a course of action, or perform any of the behaviors that research has shown to lead to cognitive dissonance? We have no firm answer to this question but it is interesting food for thought.

Of course, the Justification Hypothesis provides a clear answer as to why being unjustified is an aversive state. To effectively solve the problem of social justification, we must have a justifiable narrative of our actions.

## **The Justification Hypothesis, Attributions, and the Self-Serving Bias**

Attributions are the causal explanations people develop for why things happen. From the perspective of the Justification Hypothesis, attributions are a particular and important kind of justification. Attributions are crucial because they allow us to represent and frame the forces working in the world, both physical and social. In a subsequent section of this chapter, Leigh Shaffer's analysis of animistic attribution and its relationship to religion through the lens of the Justification Hypothesis are examined. Animistic attribution is the process by which people make purposeful attributions regarding physical events, such as when Pat Robertson attributed the 2010 earthquake in Haiti being caused by their having made a pact with the devil (James, 2010).

Our focus here is on how attributions about the self should be organized according to the Justification Hypothesis. That is, if our self-consciousness system functions to solve the problem of social justification, what implications does that insight have for how people explain what they do and the causes of their actions? One implication is that it suggests that people should generally construct their justifications in accordance with social influence, meaning that in cases where the cause is ambiguous, people should tend to give themselves the benefit of the doubt. Specifically, they should explain good things that result from their actions in a manner different than they explain bad things. Consistent with the Justification Hypothesis, people tend to explain bad outcomes in terms of external, temporal, and local causes and good outcomes in terms of internal, stable, and general causes. To be concrete about it, after failing a math test an individual is more likely to claim it was because the teacher asked unfair questions or because they were tired. However, when an individual acs the literature exam, they are more likely to think it was a

good and fair test, they studied hard, and that the result is to be expected because they are bright.

It is called the self-serving bias for obvious reasons, and the tendency has been confirmed in literally hundreds of psychological studies and can be considered one of the most robust findings in social psychology. To demonstrate the pervasiveness of this tendency, Friedrich (1996) found that after students were taught about the self-serving bias they tended to see themselves as less self-serving than most other people, a phenomenon he humorously coined the ultimate self-serving bias. The Justification Hypothesis directly accounts for these findings because people want relationships with more skillful, giving, powerful people, the more positive picture of yourself you can justifiably paint the better.

A review and meta-analysis by Mezulis, Abramson, Hyde, and Hankin (2004) found strong support for the pervasiveness of the self-serving bias. The average effect size (a metric researchers use to compare differences) across several hundred studies was large. Importantly, these authors looked at cultural differences. Although people in all cultures engage in self-serving biases, those in Western, individualistic cultures exhibited a substantially greater tendency to engage in self-serving biases than those in Eastern, communal cultures. This raises an important point to be made about the Justification Hypothesis. In introducing the idea the initial emphasis was placed on how the evolution of language must have created the problem of social justification, and that in turn shaped the adaptive design of the human self-consciousness system. The point of the Justification Hypothesis is that the self-consciousness system is designed so that it allows the individual to “download” the justification narratives of the current cultural context and utilize those narratives to navigate the social environment. Obviously, then, the justification system is sensitive to cultural context. Being born into a cultural context that emphasizes the group or collective as the primary unit of functioning will lead to a different justification system than being born into a group that emphasizes the level of the individual. And it thus follows that we would expect to see larger self-serving biases in individualistic cultures. The point here is that the Justification Hypothesis emphasizes the self-consciousness system as an organ of culture, one that is profoundly shaped by experience and cultural context.

## **The Justification Hypothesis and the Forces That Guide the Organization of Self-Knowledge**

If the self-consciousness system was shaped in response to the problem of social justification, then the way knowledge about the self is constructed and organized should be related to the kinds of pressures that would result in the capacity to respond to social justification. Flipping this on its head, we can ask: What would make someone unjustifiable in the eyes of others? Some ideas that immediately come to mind are (1) that someone has done something unacceptable (e.g., “You should not have done that!”) or that someone is undesirable (e.g., “She is ugly,” “You are stupid”); (2) or that someone has claimed something inaccurate (e.g., “That is not true,” “You don’t



know what you are talking about”); or (3) someone is inconsistent (e.g., “You said this and now you are saying that!?! You can’t have it both ways!”). Interestingly, in his review of the forces that drive the organization of self-knowledge, Brown (1997) identified three central motives: (1) the enhancement motive; (2) the accuracy motive; and (3) the consistency motive. The nature and evidence for each is reviewed below.

### ***The Enhancement Motive***

The enhancement motive is characterized by researchers as the push people experience to think and feel about themselves positively. The strong evidence for the self-serving bias reviewed above can be interpreted through the enhancement motive. That is, because people think about themselves causing good things, they feel better and are more likely to gravitate toward such explanations, with the inverse also being true. But the tendency to enhance the self is apparent in more than just attributions. In an article titled *The Totalitarian Ego*, Greenwald (1980) surveyed the vast social psychological literature on how information about the self is processed. He likened the human ego to a personal historian that is totalitarian and relentless in the manner in which it revises and fabricates history to make the individual seem more important, altruistic, and effective than the evidence would warrant. In characterizing the “positivity bias,” Brown (1997, p. 62) notes, “When it comes to their ideas about socially valued qualities and abilities (e.g., their kindness, attractiveness, and intelligence), many (if not most) do not have entirely accurate views of themselves. They regard themselves as better than they really are.” Brown’s note about the kinds of things people enhance is very important—general qualities that are socially valued. This, of course, connects directly to the Justification Hypothesis and its emphasis on the problem of social justification and how social pressures influence justification systems.

### ***The Accuracy Motive***

Although people in general tend to be biased in the way they think about themselves, they are not completely blind to reality, and there are many circumstances and reasons that people engage in what Brown calls the accuracy motive. The motive for accuracy is not hard to understand from either a behavioral investment or Justification Hypothesis perspective. If we think about our perceptions and thoughts as representations that function as maps of the outside world and then ask the question: “Does it matter if your map is accurate?” The answer is obvious. As anyone who has operated from a faulty map or faulty set of perceptions or assumptions can tell you, it can be a significant problem when we have misinformation about reality in general and ourselves in particular.

The enhancement motive and accuracy motive often exist in tension with one another, the reality of which I have seen as a clinician and experienced personally.

For example, early in my high school years, the enhancement bias was operating as I considered my “mate value.” Overestimations of my own attractiveness led me to ask out several of my female peers who were—to be blunt—out of my league in terms of physical appearance. The predictable consequence was that I was rejected on several occasions. After several such disappointments, I distinctly recall a conversation with a female friend where I was honestly and earnestly asking her for a “true” assessment of my attractiveness. Consistent with this anecdote, evidence suggests people seek accurate information to reduce uncertainty, to achieve goals, and to understand prior errors (Trope, 1986). Although I do not know of research in this area, a prediction from the Justification Hypothesis is that the accuracy motive is the motive people are most conscious of when developing their justifications. That is, when asked about what was influencing their reasoning, people will most often claim that they are trying to be as accurate as possible, as opposed to self-enhancing or consistent, which is the other motive that drives the organization of self-knowledge.

### *The Consistency Motive*

If one of the primary functions of the self-consciousness system is to generate a justification narrative that helps the individual function in and navigate the social world, it follows that the system would require a degree of stability and consistency, both for the individual and in relationship to the expectations of others. Several authors have argued that a need for consistency is a central principle of mental organization in general, and the understanding of the self in particular. As an early proponent of this view, Lecky (1945; cited in Brown, 1997, p. 52) argued the following:

According to self-consistency, the mind is a unit, an organized system of ideas. All of the ideas which belong to the system must seem to be consistent with one another. The center of the nucleus of the mind is the individual's idea or conception of himself. If a new idea seems to be consistent with the . . . individual's conception of himself, it is accepted and assimilated easily. If it seems to be inconsistent, however, it meets with resistance and is likely rejected. (p. 246)

From the vantage point of the unified theory, consistency would play a more central and organizing role in the justification system than in the behavioral investment system because of the issue of social justification.

The Justification Hypothesis predicts that people should be motivated by consistency for at least two reasons. On the one hand, as suggested by the quote above, the more inconsistency, the greater likelihood for logical contradictions and incoherence leading to potential breakdowns in the justification system. Swann (1990, 1996) has marshaled an impressive array of findings that people strive for consistent self-knowledge and seek information that verifies their existing self-knowledge. Second, although we certainly sometimes want people to change their beliefs, it is also the case that we rely on the predictability of others and the coherence of their justification systems. As a consequence, there should be social pressures for consistency, and individuals who frequently alter their beliefs, waffle in what they say,

or are flat out contradictory should be criticized, disliked, or lose social influence. Although I could not find any direct research supporting the idea that people who tend to be inconsistent are less well liked, in anecdotal support of this idea, flip-flopping is a common criticism that political opponents use against one another. For example, George W. Bush used the criticism with great effect against John Kerry in the 2004 presidential election, especially after Kerry, when speaking on his attitude for funding the Iraq war, told a crowd “I voted for the bill before I voted against it.”

## **The Justification Hypothesis and Research on Implicit and Explicit Attitudes**

An attitude is an evaluative judgment made about an object, event, or idea that is either favorable or unfavorable. Attitudes are some of the most researched constructs in cognitive and social psychology. Early definitions of attitudes generally did not discriminate between attitudes that were conscious versus those that were subconscious, presumably because of the long scholarly tradition of researchers staying away from the murky concept of consciousness (Greenwald & Banaji, 1995). But by the 1990s researchers had clearly documented the need for distinguishing between two broad domains (or kinds) of attitudes. On the one hand there are attitudes of which the individual is conscious, meaning that they can state or report on a questionnaire, and are able to offer reasons explaining the attitude. On the other hand, there are attitudes that exist outside of self-conscious awareness. Researchers initially documented the existence of a second class of attitudes by demonstrating that people could form attitudes subliminally. For example, researchers would expose subjects to a flash of an image that happened so fast they did not even that know they saw anything. However, if in the image a neutral object was subliminally paired with a noxious object, the individual would develop an unfavorable attitude toward the neutral object. This attitude would exist despite the individual being completely unaware that he or she had seen it previously (Greenwald, Klinger, & Lui, 1989). Researchers have variously called these two classes of attitudes unconscious–conscious, unaware–aware, automatic–controlled, direct–indirect, and more recently associative–propositional (Gawronski & Bodenhausen, 2006). Here I refer to the two kinds as implicit and explicit, which is consistent with most of the literature.

Although only a few decades old, the literature on the relationship between implicit and explicit attitudes is vast, and I only review a small slice of it here. First, it is useful to note that the identification of two broad classes of attitudes, one that functions on associations and is automatic, fast, and affective, and a second class of attitudes that functions on language and is deliberate, slower, and reason-based is very consistent with the two domain model of the human mind suggested by the combination of Behavioral Investment Theory and the Justification Hypothesis. But we can go further than that because the Justification Hypothesis suggests that there will be a particular kind of relationship between implicit and explicit attitudes. Given that explicit attitudes connect the individual to the social world via the

dimension of justification, we can make the prediction that where we see explicit attitudes diverging from implicit attitudes, a central variable should be the social justifiability of the attitude. That is, our explicit attitudes should be more in accordance with our identity and that which what is socially justifiable, whereas our implicit attitudes may diverge from that narrative.

One line of strong evidence for this prediction comes from examining implicit and explicit attitudes regarding race, especially the attitudes that Whites have toward Blacks in America. Dovidio and colleagues used a variety of different procedures to assess the implicit racial attitudes of Whites toward Blacks (e.g., Dovidio & Gaertner, 1993). For example, these researchers found using subliminally presented schematic faces of Blacks and Whites as primes, that White participants have faster response times to negative traits after Black than White primes and faster response times to positive traits after White than Black primes (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997). Other researchers have confirmed these findings with similar, but different response latency techniques, such as the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998).

Consistent with the framework provided by the Justification Hypothesis, the negative implicit attitudes were often divergent from the explicit attitudes, which often were positive and egalitarian. Wilson, Lindsey, and Schooler (2000) offered an explanation of these findings in a manner very consistent with the Justification Hypothesis. They argued that “dual attitudes” (i.e., when there is divergence between implicit and explicit attitudes) arise developmentally. Because Whites are exposed to negative images of Blacks through the media and to stereotypes about Blacks through common socialization experiences, they may initially develop largely negative attitudes toward Blacks. Yet they are often simultaneously given social prescriptions to be egalitarian, along with dictates about the general evils of prejudice. It is these mixed messages that give rise to dual attitudes and implicit racism. Dovidio, Gaertner, Kawakami, and Hodson (2002) offer a fascinating discussion of how such dual attitudes can set up vicious cycles of miscommunication, misunderstanding and difficulties empathizing between Whites and Blacks.

## **The Justification Hypothesis and Research on Reason Giving**

Perhaps the area that most directly connects to the Justification Hypothesis is the work done on reason giving, the process by which people give reasons for what they do and why things happen. In everyday language, reason giving and justifying are essentially synonymous. One fairly prominent class of defensive reason giving is excuses, which were defined by Snyder, Higgins, and Stucky (1983) as explanations that are designed to lessen the negative implication of an actor’s performance, thereby maintaining a positive image for oneself and others. These authors argue that excuses are prevalent in social exchanges and are used regularly when things are not going well and the individual or group or institution is threatened with a negative evaluation or consequence.

The authors offer a straightforward taxonomy of excuses as consisting of three broad classes that include denying the individual had any responsibility (the “I did not do it” class), minimizing the negative consequences (the “It wasn’t so bad” class), and acknowledging but redirecting (the “Yes, but . . .” class). Interestingly, the authors begin their survey with a chapter on the philosophical implications of excuse making, which they argue relates to debates about free will, determinism, and personal responsibility. Their theory of why people make excuses is very consonant with the Justification Hypothesis and the Influence Matrix. Simply put, the authors argue the reasons people give for their behavior have social consequences, and people are motivated to maintain a positive image of themselves (private self-concept) and of the way others view them (public self). Thus when situations arise that might threaten the social capital in either the eyes of others or their own, people generate reasons to minimize, deflect, or defend against the loss of social capital from occurring.

In his book *Explaining and Arguing: The Social Organization of Accounts*, Antaki (1994) takes an even broader look at reason giving. Antaki reviews the way a number of different perspectives in the social sciences have attempted to tackle the pervasive tendency of humans to engage in explaining and arguing in social exchange. He highlights the different approaches to the topic taken, which range from cognitive psychologists working from an attribution theory framework who attempt to understand the cognitive processes underlying the generation of explanations to micro-level sociologists who analyze the function conversation has on regulating the actions of individuals. Antaki’s book is scholarly and is difficult to summarize tersely, so readers are referred directly to it for a sophisticated overview of the various approaches to analyzing the process by which humans give accounts.

A more generally accessible account of reason giving was recently offered in the book *Why? What Happens When People Give Reasons. . . and Why* by Charles Tilly (2008). Tilly opens the book with a fairly dramatic review of what happened to people as the events of September 11, 2001 unfolded. Specifically, people everywhere were caught completely off guard by the tragic events, and this resulted in a frantic search for the reasons for what was happening. By using this example, Tilly highlights an important point about reasoning and events. Humans are always operating at least implicitly from a justification narrative for what is happening and why. When events unfold that are in contrast to that implicit narrative, the need to make sense out of them and to have a shared justification narrative with others for what is happening is extremely powerful. The more dramatic and unexpected the events, the greater the need to for such a narrative, a point Tilly makes clear in his review of the way people responded to the September 11th attacks. Moreover, Tilly (2008, p. 8) sees the centrality of reason giving in human behavior. He wrote

As eyewitnesses at the World Trade Center and Pentagon searched for reasons, they followed an extremely general human routine. We might even define human beings as reason-giving animals. While, by some definitions, other primates employ language, tools and even culture, only humans start offering and demanding reasons while young, then continue through life looking for reasons why.

Tilly's (2008) book is built on an interesting taxonomy of reason giving. Specifically, he argued that there are four broad categories or contexts that frame reason giving:

1. *Conventions* are reasons given in social exchanges that frame everyday actions, usually offering an explanation for an action or related event that was unexpected or potentially negative, such as excuses for being late (e.g., the traffic was bad), explanations for success or achievement (e.g., I worked really hard on this), or reasons that justify deviations from the norm (e.g., he beats to his own drummer).
2. *Stories* are more detailed narratives that explain longer cause-effect sequences, such as the events of 9/11. They are generally shared in more intimate relationships, in contrast to conventions which are given between acquaintances. Stories themselves are usually relational in nature, having protagonists, plots, antagonists, and ending. Individuals place their broader actions in storied accounts and share those accounts with important others. Thus, individuals will share stories of their first love, when they failed at a big event, or why their family moved.
3. *Codes* are governing policies that officially regulate and legitimize practices. Thus, laws and legal procedures, company policies, ethical codes, and religious decrees are all codes and people often justify action or its inhibition on the basis of codes.
4. *Technical accounts* are the specialized, often scientific causal explanations of events. They are given by experts often in a language that is not accessible to lay people.

Although there are likely many different approaches to developing taxonomies of justifications, Tilly's four contextual categories seem reasonable to me, and I have generally found that I can usually classify explicit justifications as falling into one or the other category, although as Tilly himself admits, they blend into one another quite frequently. One noteworthy aspect of his categories is that they stretch from justifications that are more social on the one hand (conventions and stories) to those that are more analytical on the other (codes and technical accounts). This raises one of the most important points about the concept of justification, which is that the concept stretches from defensive rationalizations through social conventions into explaining and arguing and finally into technical accounts and scientific explanations. Up until this point, the examples we have been discussing generally emphasize social reason giving, and it seems to paint a picture of people as rationalizers. However, as discussed in [Chapter 1](#), the argument is not that humans are rationalizers. That is too pejorative and overly simplistic. The argument is that they are justifiers. And the process of justification bridges us from social reason-giving to analytical reasoning.

## The Justification Hypothesis and Research on Reasoning

One only needs to read the lucid description of great ape behavior offered by primatologists like Franz de Waal (1982), Diane Fossey (1983), and Jane Goodall (1986) to realize that our nearest relatives live intricate and complicated social lives. These

descriptions strongly suggest major similarities between humans and other great apes in the basic biopsychological motivational and affective systems that guide behavior in the social domain. At the same time, one only needs to confront our great ape relatives with tasks that require basic analytic reasoning to realize that the gulf between humans and other great apes is oceanic in this domain. Why, according to the Justification Hypothesis, would humans be good reasoners? For the same reason that we pay smart lawyers more than foolish ones. Determining logical inconsistencies in ones' own and others' justification systems is obviously of crucial importance. The only way to identify such logical inconsistencies is via the process of analytic reasoning. Some evolutionary psychologists are fond of pointing out that there cannot be a domain general learning device because of the frame problem and because there are no general adaptive problems that must be solved (Tooby & Cosmides, 1992). The Justification Hypothesis challenges this assertion, at least in the sense that to solve the problem of what is and what is not justifiable requires some capacity for analytic reasoning. However, the Justification Hypothesis does not just provide a frame for understanding why humans have capacities to reason, it actually goes further and specifies exactly what kinds of reasoning skills we humans will have developed best.

A logical corollary of the Justification Hypothesis is that the general reasoning capacity in humans emerges out of determining what is and what is not justifiable in the social context, and this gives rise to another implication of the Justification Hypothesis. If social reasoning gave rise to general reasoning, then humans should be particularly adept at social reasoning, at least in comparison to other forms of general reasoning. This is precisely the case. Cognitive psychologists have long noted that humans reason more effectively about what one may, ought, or must not do in a given set of social circumstances than they do when reasoning generally. Cognitive psychologists refer to reasoning about socially justifiable acts as deontic reasoning. After noting how crucial it is to be able to reason about what is and is not socially justifiable in virtually all social situations, Cummins (1996a, p. 823) summarized the findings on deontic reasoning in adults as follows:

In contrast to their performance on statistical reasoning (e.g., Kahneman, Slovic, & Tversky, 1982), indicative reasoning (e.g., Wason & Johnson-Laird, 1972), and mathematical or scientific problem-solving tasks (e.g., Chi, Feltovich, & Glaser, 1981), adults typically perform consistently and well on tasks requiring deontic reasoning (e.g., Cheng & Holyoak, 1985, 1989; Griggs & Cox, 1983; Manktelow & Over, 1991, 1995). In fact, so robust and reliable is performance on deontic tasks that numerous proposals have been put forth to explain it.

Cummins (1996a) proceeded to demonstrate that 3- and 4-year-old children also show superiority when reasoning about what is and what is not socially justifiable compared to tasks requiring general reasoning. In a separate article arguing that the ability for deontic reasoning is a consequence of evolutionary pressures, Cummins (1996b) observed that deontic reasoning “emerges early in childhood, is observed regardless of the cultural background of the reasoner, and can be selectively disrupted at the neurological level” (p. 160). In short and in direct accordance with the Justification Hypothesis, there is an abundance of evidence that suggests that humans reason better about what is and what is not socially justifiable than they do when reasoning about abstract general truths.

The Justification Hypothesis further characterizes human reasoning as ultimately emerging out of the behavioral investment system and being intimately tied to social influence processes and this gives rise to a different picture regarding the relationship between motives and reasoning than a pure reasoning model. A pure reasoning model of human higher thought, advanced by some researchers (see, e.g., Miller & Ross, 1975), posits that the language-based cognitive system uses logic and evidence to arrive at conclusions that are as accurate as possible. In such a model, motives or goals follow after reasoned conclusions are reached. Spock from Star Trek or the notion of a dispassionate scientist in the lab crunching objective data represents the analytic reasoning ideal.

The Justification Hypothesis, especially in combination with Behavioral Investment Theory and the Influence Matrix, argues for a different view of reasoning. Instead of a disconnected, pure reasoning system that processes information according to formal rules of logic, the view afforded by the unified theory is that of a motivated reasoning system (e.g., Kunda, 1990). Motivated reasoning argues that more often than not we use our reasoning process to arrive at previously determined desired outcomes. Thus, in motivated reasoning, the goal or desired outcome generally comes first and the reasoning about the goal follows. Although under some conditions our primary desire is to arrive at an accurate conclusion (for example, when we are confused or uncertain or when receiving a reward for arriving at the correct answers), much of the time—especially in day to day social interaction—individuals have desired outcomes that they want to arrive at first and then reason back from that goal. So according to a motivated reasoning model, instead of usually relying on pure reasoning to arrive at goals, the more common everyday sequence is that individuals have desired outcomes that they want to move toward first “but their ability to do so is constrained by their ability to construct seemingly reasonable justifications for these conclusions” (Kunda, 1990, p. 480). In this view the human reasoning system is much more akin to that of a defense attorney whose reasoning is framed by a predetermined goal state than it is to an independent, dispassionate scientist attempting to decipher the objective truth.

Haidt (2001) offers an analysis of moral reasoning and judgment that is very congruent with the model of human cognitive processes suggested by the unified theory. Called the social intuitionist approach, Haidt argued that pure moral reasoning rarely causes moral judgment. Instead, Haidt posited that humans have both intuitive (experiential) and rational (justifying) systems of cognition, and that they way most people arrive at moral judgments is they arrive at a sense as to whether the situation in question was good or bad through the intuitive system, and then utilize their justification system to generate post hoc rationalizations as to why. Specifically, “People have quick and automatic moral intuitions, and when called on to justify these intuitions they generate post hoc justifications out of a priori moral theories” (Haidt, 2001, p. 823). Haidt argued that two major classes of motives direct and guide reasoning. First and foremost are relatedness motives, which translate roughly into the various relational dimensions highlighted by the Influence Matrix, and includes concerns about impression management and smooth interactions with



people. Second, people have motives to maintain consistent and coherent, so that their overall narrative is justifiable.

In this section we have reviewed a wide variety of different domains of research that can be integrated and assimilated through the lens of the Justification Hypothesis. We have shown that the formulation that the self-consciousness system evolved in response to the adaptive problem of social justification leads to a picture of that system that can readily account for findings in cognitive, developmental, and social psychology. Specifically, the Justification Hypothesis provides a causal explanatory framework for the interpreter function of the left hemisphere, the forces that organize self-knowledge, self-serving biases, cognitive dissonance, explicit and implicit attitudes, social reason giving, and reasoning in general. This shows that the Justification Hypothesis is a powerful lens for human psychology. However, one of the great advantages of the Justification Hypothesis is that it not only frames insights for how individuals behave, but it is a consonant framework for societal level processes. In this next section, I review how the Justification Hypothesis can frame the evolution of human culture and why we can understand the defining feature of human culture as being the emergence of large-scale collective justification systems.

## **Culture As the Rise of Collective Justification Systems**

Anatomically modern humans existed more than 100,000 years ago but with crude technology, little or no art, and primitive social interaction. However, by 50,000 years ago a pattern of cumulative growth began that would result in a creative, technological, and social explosion. During the past 10,000 years, the pace of innovation has continued, including such achievements as agriculture, specialized division of labor, and systems of written language. The nineteenth and twentieth centuries saw a marked continuation of knowledge and technological growth, a pattern Piel (1972) characterized as the acceleration of history. The rate and character of such change makes it clear that explanations for it will not be found simply in terms of biological evolution through natural selection but instead must be considered to be a consequence of socio-cultural evolution.

Although culture has been notoriously difficult to define, theorists now generally agree that the term “culture” does not refer to the totality of human lived experience or all aspects of society, but it specifically refers to the patterned sphere of beliefs, values, symbols, signs and discourses that are shared by a group of people (Smith, 2001). As framed by the Justification Hypothesis and the ToK System, Culture (with a capital “C”) is the fourth dimension of complexity, and consists of symbolically mediated justification systems that exist at the large-scale level and function to frame human action. Culture, then, is theoretically separable from spheres of technology, human behavioral investment patterns, and the biophysical ecology in which humans live, although there are clearly complex interrelations between these various domains. With Culture so defined, we can then ask, “Where did Culture come from?” and “Why is the rate of cultural evolution accelerating?”

The Justification Hypothesis answers the question, “Where does Culture come from?” by locating the combustible spark as the problem of social justification that emerged as language evolved enough sophistication to allow for a window into the minds of others via the capacity to ask questions. The result of this spark was the emergence of justification systems at the individual and small group level and these systems then become networked together to coordinate populations of people. This section elaborates on arguments for how the Justification Hypothesis frames Culture. First, I briefly describe how the Justification Hypothesis provides a frame for the cultural transmission of technology through justification systems that prescribe procedural knowledge and actions. Second, I touch on how social processes lead to the emergence of norms and how these norms impact human behavior. Third, the argument is made that justification systems carry with them the necessary ingredients to spawn their own evolution. And, finally, I show how the Justification Hypothesis provides a framework for understanding societal institutions like religion, law, and science, and why such systems have become increasingly complex and differentiated.

## **The Justification Hypothesis and the Transmission of Technology**

The sophistication of the tools and artifacts developed by modern humans (*Homo sapiens*) was far greater than those developed by Neanderthals (*Homo neanderthalensis*), suggesting more advanced cognitive capacities (Wynn & Coolidge, 2008). Moreover, unlike the Neanderthals whose technological sophistication remained fairly constant over a 200,000 year period, in the last 50,000 years the technology of modern humans evolved quite rapidly. Shaffer (1981, 1998) has articulated a conceptualization that links justification processes and transmission with the evolution of technology. He argued that cultural skills and technological developments are transmitted via packets of “recipe knowledge,” which are justified by knowledgeable experts to novices. To envision how this process occurs, imagine a circumstance where novices observe an experienced toolmaker and form “why” questions: “Why do you strike the stone at this angle?” or “Why do you carve the bone this way?” A likely response given by an expert toolmaker would have been a justification: “I strike the rock this way to make it flake—if you strike it that way, the hammer will glance off the rock and you will end up striking your hand!” or “If you carve the bone this way it will crack and the hook will be useless.”

This analysis suggests that processes of justification will lead to systems of reasoning that will be amenable to planning and innovation. And in contrast to Neanderthals, the technology of early humans was indicative of such planning. For example, as early as 10,000 years ago, humans learned to build fish weirs across streams to trap fish within easy reach of the shorelines (Wynn & Coolidge, 2008) and to build kites by gathering and piling large rocks at narrow places in the paths of migratory animals or the entrances to valleys in order to funnel and corral fleeing animals to fall into pits and traps close to hunters’ blinds and to bring them within

easier reach of spears or arrows (Betts, 1987). These weirs and kites required considerable time and effort to construct, and thus hunters had to forego opportunities for traditional hunting activities during the construction process. Importantly, this potentially costly innovation requires justification: “If we have always done this, why should we do *that* instead?” The social influence needed to coordinate the construction of such structures must have required leaders to justify the expending of time and energy. In sum, justification processes provide a readily accessible way to tool construction and tool use to justification via recipes and legitimized purpose.

## The Justification Hypothesis and the Emergence of Norms

A second explanation for the emergence and accelerating rate of cultural evolution provided by the Justification Hypothesis is the fact that sociolinguistic exchanges between individuals give rise to complex systems of justification (Shaffer, 2005). This process was examined empirically in the influential work of Muzafer Sherif. Sherif realized that the apparent movement of a stationary point of light in a dark room (the autokinetic effect) represented a means to create a laboratory analog of events in early human history before the establishment of shared norms, which Sherif believed was the essence of culture. Sherif (1966) found that, when participants were asked to make judgments of the apparent motion by themselves, their initial judgments usually exhibited considerable variability. With repeated judgments, participants reduced the variability around a personal anchor termed a “personal norm.” Others’ evaluations, however, were typically very influential in the participants’ subsequent judgments. Moreover, eventually groups would establish norms that would specify the legitimate perception of movement in a manner that carried moral overtones.

The powerful influence that shared norms can have on individuals was demonstrated in an equally influential study of social conformity by Solomon Asch (1956). As is now well known, participants in this study were asked to make perceptual judgments regarding the lengths of lines, and although the answers were clear, participants were asked to give their answers in a group format following confederates who would give incorrect judgments to certain responses. Although the strength of the confederates’ influence varied with their number and degree of unanimity, a substantial portion of participants gave what they knew to be incorrect judgments in the interest of conforming. Consistent with the Justification Hypothesis, these individuals reported being concerned about the social justifiability of their responses. For example, Asch quotes one subject as reporting that, “You have the idea that the attention of the group is focused on you. I didn’t want to seem different. I didn’t want to seem an imbecile” (Asch, 1956; p. 31). In short, Asch’s elaboration of Sherif’s paradigm demonstrates how actors feel social pressures expected under the Justification Hypothesis. To perform in concert with others is to perform in a justified fashion, but to perform in violation of group standards is to experience the discomfort associated with adopting an unjustifiable stance. Taken together, these classic studies of social influence illuminate the omnipresent selective pressure that

leads to the development of justification systems composed of norms that can affect a whole group.

## **The Justification Hypothesis and the Emergence of the Cultural Dimension of Complexity**

We have reviewed how the Justification Hypothesis provides a framework for the cultural transmission of technology through justification systems that prescribe procedural knowledge and actions and how the idea is consistent with research on how norms emerge in groups and how these norms impact human behavior. A third advantage of the model of Culture afforded by the Justification Hypothesis is that it provides a readily accessible frame for understanding why it evolves. Justifications are a great example of what Dawkins (1989) called a meme, which is a unit of cultural evolution. Much like genes, justifications interlock to form complex, functional systemic networks. And such systems can easily be envisioned to evolve. To do so, consider how justifications offered by individuals sometimes become accepted by the group, and thereby come to provide the informational “glue” necessary to coordinate the behavior of large groups of people. Confidence that justification systems would evolve is enhanced by the fact that they clearly involve the three key elements of evolution: variation (different justifications are offered), selection (certain justifications are better at legitimizing action than others) and retention (selected justifications are stored and repeated). Kuhn’s (1996) analysis of scientific revolutions is obviously commensurate with the idea that justification systems evolve. In Kuhn’s view, scientific paradigms—i.e., scientific justification systems—begin to weaken when data cannot be accounted for by the existing explanatory system. Such data are used to show the current system is unjustifiable, and adherents of new scientific justification systems will emerge.

## **Culture as Large-Scale Collective Systems of Justification**

The final way in which the Justification Hypothesis provides a framework for understanding Culture is that it provides a framework for understanding human beliefs and values at the macro-level. Science, law, and religion can all be readily considered to be large-scale collective justification systems. One question the Justification Hypothesis immediately helps to address is why there is such a diversity of different kinds of justification systems. Traditional societies are relatively undifferentiated, with religious worldviews, explanations for natural phenomena, and prescriptions for social conduct provided by global justification narratives. In contrast modern developed nations have distinct systems of justification, with religion separated from law, government, philosophy, science, and other cultural institutions. The reason, according to the framework afforded by the Justification Hypothesis, is because different foundational goals serve as a force for differentiation within the larger stream of cultural evolution. We have both science and religion in modern

society because humans have needs for describing the world around them as accurately as possible, and humans also have needs for determining right from wrong and finding meaning and purpose in the world in which they live.

Shaffer (2008) argued strongly for the utility of conceptualizing religion as large-scale justification systems. He notes

While the term “religion” symbolizes a staggeringly diverse set of individual systems, . . . many documented practices and beliefs are clearly related to the process of finding justification both in human conduct but also in the “actions” of supernatural forces, spirits, or gods. Religions not only develop normative systems that define a believer’s obligations and serve as justification systems for communities of like-minded believers, but religions often develop belief systems (that sociologists call “theodices”) that serve to justify the actions of God himself!

Shaffer (2008) argued persuasively that the framework afforded by the Justification Hypothesis explains such diverse features of religious phenomena as (1) the normative function religious beliefs play in legitimizing actions; (2) the nature of *theodices*, which are the belief systems that serve to justify the actions of God(s); and (3) animistic attribution, which is the process by which people come to see natural events as being caused by personal intent.

The following story makes these elements clear.<sup>1</sup> In 1984, Anglican priest David Jenkins had been nominated for the post of Bishop of Durham. Jenkins had an impressive record of church service and was considered well-qualified for the post except for one problem: Jenkins had publicly stated that he did not believe in the literal interpretation of the doctrines of the Resurrection or the virgin birth of Jesus Christ (Iyer & White, 1984). Despite cries from its evangelical wing, the Anglican Church had decided to confirm his appointment and to consummate the process with a ceremony in the historic cathedral at York, Minster. The ceremony was held on a Sunday. The following Wednesday evening during a severe thunderstorm, the wooden roof of the south transept of the church was struck by lightning and caught fire. Following a night of battling the flames, the fire was contained to the wing where the bolt originally struck, but even before the flames were extinguished the Anglicans had already begun to debate the “angry God” hypothesis. Not surprisingly, the battle lines were drawn for this debate along the same lines that had been drawn over the Bishop’s confirmation in the first place—the two sides corresponded roughly to those who held traditional, evangelical views and those who held to more recent, modernist views. Time magazine captured the clash of worldviews this way: “It was an unholy coincidence that many took to be divine retribution” (“A bolt from the heavens,” 1984, p. 62).

It seems clear that the Justification Hypothesis readily accounts for much of this basic narrative: Jenkins was perceived to represent a threat to the cohesiveness of the justification systems of the evangelical wing and, hence, these individuals wove a series of events into an account that allowed them to claim legitimacy for their position of resisting his appointment. However, there is an aspect of this narrative

---

<sup>1</sup>I am indebted to Leigh Shaffer for this story and analysis.

that seems at first blush to be somewhat distally removed from the Justification Hypothesis: the process by which these individuals came to see the lightning strikes as personally caused misfortunes and sought reasons rather than causes for these events. In other words why did these individuals make purposeful attributions about ostensibly inanimate occurrences? Shaffer (2008), who offers a detailed analysis of both animistic attribution and theodices, answers this question as follows:

But why would the [Justification Hypothesis] tend to favor individuals who see the universe in anthropomorphic terms and to favor societies that would develop formal systems of religious belief, ritual, and moral codes? My answer is that just as a species that had developed a Theory of Mind would question the justifiability of others' actions and develop a looking-glass self to justify its own actions, such a species would also turn its "mind-reading" capacities toward natural phenomena, begin to make animistic attributions and elaborate them into depictions of the world of the gods in anthropomorphic terms, and then try to find ways to justify its conduct to the gods and try to hold the very gods of the universe accountable for their actions.

## Conclusion

As I hope is clear from the above analyses, with the concept of justification systems, the Justification Hypothesis has the capacity to link the individual with social levels of analyses, thus providing a frame to bridge psychology with the rest of the social sciences. Moreover, when considered in concert with Behavioral Investment Theory and the Influence Matrix, the Justification Hypothesis is directly bridging an important gap in the current knowledge landscape, the gap between the natural and social sciences. Edward O. Wilson, whose ideas and contributions are explored in detail in the next chapter, explicitly characterized the rift between the natural and social sciences as one of the great remaining problems in science. He further suggested that the lack of a coherent framework that integrates the natural and social sciences prevents the science of human behavior from effectively progressing. He wrote

We know that virtually all of human behavior is transmitted by culture. We also know that biology has an important effect on the origin of culture and its transmission. The question remaining is how biology and culture interact, and in particular how they interact across all societies to create the commonalities of human nature. What, in the final analysis, joins the deep, mostly genetic history of the species as a whole to the more recent cultural histories of far-flung societies? That, in my opinion, is the nub of the relationship between the two cultures. It can be stated as a problem to be solved, the central problem of the social sciences and the humanities, and simultaneously one of the great remaining problems of the natural sciences.

At present time no one has a solution. But in the sense that no one in 1842 knew the true cause of evolution and in 1952 no one knew the nature of the genetic code, the way to solve the problem may lie within our grasp. (Wilson, 1998, p. 126)

The Justification Hypothesis, in conjunction with the rest of the unified theory, provides the solution and clearly articulates the link between the natural (or perhaps more appropriately, nonhuman) and social (or human) sciences. The essence

of Culture is the presence of large-scale justification systems that function to coordinate and legitimize human behavior, and the fundamental point of a social science perspective is that human behavior must be understood in the context of the larger socio-linguistic system in which it is immersed. The theoretical problem has been that there was no systematic way to understand how the evolution of mind in general, and the human mind in particular, led to the emergence of these cultural justification systems. As such, social scientists have tended to focus simply on the systems themselves and not concern themselves with the origins of their emergence. As an inevitable consequence of this starting point, these social science models lacked any systematic framework for integrating biopsychological causation and thus were obviously incomplete. Unfortunately, however, the biopsychological models built by taking a bottom-up perspective have proven inadequate for providing a framework for understanding the emergence of the large-scale justification systems examined by macro-level social scientists. As a consequence, we have biopsychological levels of causation essentially disconnected from sociocultural levels, something we saw clearly in [Chapter 2](#) when we examined the debate regarding gender differences between evolutionary psychologists and social role theorists.

The Justification Hypothesis completely changes this state of affairs because it (1) offers a clear formulation of the evolutionary changes in mind that gave rise to human culture; (2) offers a theory of human self-consciousness that links human psychological with sociological levels of analysis; (3) integrates a wide variety of different theoretical perspectives (e.g., psychodynamic theory, social cognitive theory, everyday life sociology) into a coherent whole; (4) organizes and explains vast domains of empirical data; and (5) offers a framework of explanation that is consonant with frameworks in the social sciences.

Prior to delving into the fourth piece of the unified theory, let me offer a general summary of the terrain we have covered. The argument so far has been that the field of psychology is a thicket of conceptual problems, consisting of a multitude of conflicting and overlapping paradigms, and the nature of these problems are such that they cannot be solved with ever-increasingly sophisticated data collection methods. Instead, what the field desperately needs is a meta-theoretical perspective that allows us to integrate the key insights from each major perspective and see the elephant as a whole.

Toward that end, three large pieces of the puzzle have been offered. The first piece was Behavioral Investment Theory, which provides a general theory of animal behavior built on an evolutionary foundation that merges computational/cognitive science with behavioral and neuroscience. With Behavioral Investment Theory we can knit together the many of the branches of psychology that are most strongly associated with the natural sciences. Building off of the Behavioral Investment Theory framework, a schematic of the architecture of the human mind was presented that consisted of four levels: (1) the sensory-motor; (2) operant-experiential; (3) imaginative thought; and (4) human self-consciousness. We then introduced the Influence Matrix, which in many ways serves as a bridge between Behavioral Investment Theory and the Justification Hypothesis. On the one hand, the Influence Matrix is an outgrowth of Behavioral Investment Theory, applied to human social

motivation and emotion. On the other hand, the problem of social justification is intimately tied to social influence. With the Influence Matrix, we developed a model that allows us to richly view dynamic relational processes that guide humans and shapes the kind of justifications they offer.

In this chapter we have detailed the Justification Hypothesis, which argues that the evolution of language created the problem of social justification, and this adaptive problem spurred the evolution of both the human self-consciousness system and Culture. In conjunction with the Justification Hypothesis, we introduced a new tripartite model of human consciousness that consists of an experiential self, a private self, and a public self, along with filtering processes between these domains. We then reviewed many different domains of research, such as cognitive dissonance, attributional biases, human reasoning and reason giving and demonstrated that the Justification Hypothesis forms a hub of an idea that can assimilate and integrate many lines of research.

The final piece of the unified theory to be explored is the Tree of Knowledge System. As was briefly touched upon in the first chapter, the Tree of Knowledge System is a map of cosmic evolution that defines key concepts like Matter, Life, Mind, and Culture, and depicts how they exist in relationship to one another. The need for such a meta-theoretical framework is clear when one considers the problem of psychology. For unlike physics or biology, psychology—with its joint focus on both animal behavior in general and human behavior at the level of the individual—has confusingly spanned across two dimensions of complexity, that of Mind and Culture. In the next chapter we examine the ToK System and how it provides a way to bridge the terrain between the natural and social sciences and solve the problem of psychology.