Modeling Personality and Individual Differences: The Approach-Avoid-Conflict Triad

Karl Fua\textsuperscript{1,2}, William Revelle\textsuperscript{1} and Andrew Ortony\textsuperscript{1,2}
(karl.fua@gmail.com, {revelle,orton}@northwestern.edu)
\textsuperscript{1}Northwestern University, Evanston, IL 60208 USA
\textsuperscript{2}Computational Cognition for Social Systems, Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore

Abstract
Personality is the unique patterning of affect, behavior, cognition and desires in individuals across time and situations. This patterning can occur on different information processing levels, specifically, the Reactive, Routine, and Reflective levels (Ortony et al., 2005), across these four domains. Reinforcement Sensitivity Theory (RST; Gray & McNaughton, 2000) provides a biological account of the functional subdivision of the reactive level into the approach, avoidance, and conflict systems. These systems differ in their sensitivities to different classes of cues, giving rise to personality differences. But, individuals also differ at the routine and reflective levels in terms of how they respond (cognitively, affectively, behaviorally and motivationally) to approach situations, avoidance situations, and internal conflicts. In this paper, we discuss how the approach-avoidance-conflict (AAC) triad can be used as a broad framework for incorporating personality and individual differences into theories of human cognitive architectures. We also present work in progress on a computer implementation of the AAC triad at the reactive level.

Keywords: Reinforcement Sensitivity Theory, Personality, Behavior, Affect, Motivation.

Introduction
A trolley is hurtling along a track toward five children, all of whom are tied to the track. Should you flip a switch to divert the trolley onto another track on which only one child is tied so that only one life is sacrificed instead of five?

The moral dilemma posed by this well-known “trolley problem” illustrates the presence of high level motivational conflicts that arise when incompatible goals and values are coactivated. In this case, one would want to save the five lives but at the same time avoid taking the life of another. The recognition that the conflict exists results in rumination and reasoning as an individual seeks to resolve the conflict. Internal conflicts of this kind often occur in social situations, as for example, when a person wants to approach a potential date while also wanting to avoid rejection. Such conflicts lead to indecisive behavior such as dithering between approach and shyly looking away, and paying greater attention to hints that might inform the individual if approach (or shying away) would be a more suitable course of action.

Important in the present context is the fact that individuals differ in how they perceive and weigh alternatives (taking a life versus not saving a life), and how they handle different (approach or avoidance) goals and conflicts. Reward oriented individuals are prone to engage in riskier behavior, such as brazenly approaching a potential date. In the same situation, punishment oriented (averse) individuals will be more likely to shy away for fear of rejection. Yet others, who are prone to indecisiveness, are likely to spend time ruminating about the pros and cons of approaching and avoiding. We believe that this patterning of affect, behavior, cognition and motivation occurs across all three information processing levels proposed by Ortony et al. (2005) in their (ONR) model, from the reactive (lowest) to the reflective (highest). For instance, chronically goal/reward oriented individuals tend to exhibit more pro-social behavior (e.g., attending lively parties, dating more often; Paunonen, 2003), and are biased toward speed (maximizing hits) rather than accuracy (avoiding misses) when completing simple reactive tasks (Higgins & Spiegel, 2004).

Many models of the human cognitive architecture have been proposed, for example in cognitive psychology (Anderson & Lebiere, 1998; Ortony et al., 2005; Broadbent, 1971), personality psychology (Carver et al., 2009; Revelle, 1993), and artificial intelligence (Sloman & Chrisley, 2005; Newell, 1990). Many of these architectures, such as H-CogAff, ACT-R, and Soar, are highly elaborated and have been studied in great detail. However, although personality is a key moderator of individuals’ affect, behavior, cognition and motivations, there has been little effort to include an account of personality and individual differences in these architectures. On the other hand, computational models of personality focus on describing specific aspects (e.g., the motivational aspect; Read et al., 2010) of personality but not the systematic integration of personality in the broader framework of cognitive, affective, motivational, and behavioral processes.

In this paper, we argue that there are three main classes of sensitivities (sensitivity to cues for reward, cues for punishment, and internal conflict), related to approach, avoidance, and conflict resolution (AAC) respectively, and that the AAC framework has implications for how different cognitive processes interact with each other and how the mechanisms driving personality and individual differences can be modeled systematically. Behavioral/motivational processes are commonly studied in terms of approach and avoidance (e.g., Carver et al., 2009; Elliot & Thrash, 2002). An individual’s sensitivities to cues for reward and for punishment refer to how that individual’s approach and avoidance systems react to and learn from such cues. Inspired by RST, we propose that conflict resolution is a third component that should be considered in conjunction with the approach-avoidance pair, and that it is associated with the sensitivity to conflicts. We define sensitivity to conflicts as the threshold beyond
which incompatible behavioral tendencies activate the conflict system—a system that we take to be distinct from the approach and avoidance systems and that is responsible for triggering conflict resolution processes (e.g., information gathering and rumination). Differences in the three kinds of sensitivities underlie broad personality dimensions such as the Giant 3 (Extraversion–Neuroticism–Psychoticism; Eysenck et al., 1985) or the Big 5 (Openness, Neuroticism, Conscientiousness, Extraversion, Agreeableness; Costa & McCrae, 1992; Goldberg, 1990). The AAC triad allows us to study how personality arises in individuals and how it influences cognitive processes like strategizing or resolving dilemmas, enabling us to address questions such as why, in the same situation, the plans an extravert makes differ from those of an introvert?

We believe that a theory of human cognitive architecture should be capable of accommodating differences in the ways in which different individuals feel, want, think, and act. To this end, our current work examines the structure of the reactive, routine and reflective levels (in the ONR model) in terms of the AAC triad, and looks at how these structures influence the organization of different parameters in systems. Our goal here is to propose a general framework that augments existing architectures to help in thinking about personality, and to elucidate how high and low level processes interact with each other.

The Approach-Avoidance-Conflict (AAC) Triad: From the Reactive to the Reflective Level

The Reactive Level

According to the ONR model, at the lowest level, behavioral responses to environmental cues are immediate and reactive. Automatic responses like the instinctive flight at the sight of a predator belong to this level. Reinforcement Sensitivity Theory (RST; Gray & McNaughton, 2000) was originally developed as an animal model of fear and anxiety, and has also been extensively studied in personality psychology (Corr, 2008; Smillie et al., 2010). RST proposes three functionally distinct subsystems—the Behavioral Approach System (BAS), the Fight-Flight-Freeze System (FFFS) and the Behavioral Inhibition System (BIS), each responding to different classes of cues with different sensitivities. RST offers neurobiological evidence that low-level, rapid behavioral responses, which we think of as the reactive level, have, at least functionally, the approach-avoid-conflict triadic structure.

The Approach-Avoid-Conflict Triad in RST  The BAS, FFFS and BIS handle approach, avoidance, and conflict respectively. The approach system (BAS) is associated with the dopamine system and reacts to cues for reward, and is implicated in the learning of reinforcing signals of reward. The reactivity of BAS is highly correlated with trait extraversion and an individual who has an overactive BAS is prone to exhibit impulsive approach behaviors toward hedonic rewards. Similarly, the avoidance system (FFFS) handles cues for punishment. The FFFS is primarily associated with fear, panic and avoidance behaviors, resulting from the activation of the periaqueductal grey, medial hypothalamus and amygdala regions of the neural system. The avoidance system is specifically modulated by panicolytic (suppression) and panicogenic (stimulating) drugs; individuals with a high sensitivity to cues for punishment are susceptible to phobias and panic attacks.

A major part of RST focuses on the functions of the conflict system (BIS). This system is associated with the septo-hippocampal system (SHS) and its major role is to detect conflicts and trigger appropriate conflict resolution behavior. The BIS handles two forms of conflicts: conflicts in motivations, and conflicts in expectations. Motivation conflicts occur between or within the approach and avoidance systems. An example of an approach-avoid conflict is the desire to escape from a burning building conflicting with the desire to save a trapped loved one. Expectation conflicts occur either when a stimulus is detected but not expected (novelty) or when an expected stimulus is absent. Examples would be suddenly seeing a furtive shadow in your house at night (novelty), or turning on the lights expecting to see a burglar but seeing an empty room instead (absent expected stimulus). The BIS detects such expectation violations with a comparator (the CA3-comparator in the SHS) that compares the signal (presence or absence of an expectation) from the entorhinal cortex stream and the signal (presence or absence of an actual stimulus) from the medial septum. When conflicts are detected, the BIS sends inhibiting signals to the conflicting systems to inhibit prepotent responses, and triggers behaviors such as information gathering. Importantly, the BIS does not actually resolve conflicts but rather triggers potentially appropriate higher-level cognitive processes and behaviors to do the resolution. Unlike the FFFS, the BIS has been shown to be insensitive to panicolytics/panicogensics but instead responds to anxiolytic/anxiogenic drugs. The BIS is therefore a separate system that is specifically associated with anxiety (as opposed to fear that is associated with the FFFS) and is implicated in Generalized Anxiety Disorders. The BIS is also highly correlated with trait anxiety and neuroticism. At least from a functional standpoint, different individuals must possess different thresholds (sensitivities to conflict) for the activation of the conflict system, and these thresholds are independent of and exist in parallel with the sensitivities to cues for reward and punishment that reside in the approach and avoidance systems respectively.

The Routine Level

The routine level resides between the reactive and reflective levels and deals with habitual, routinized behaviors. It deals with expectations over a longer time span than the moment-to-moment activities in the reactive level. While the reactive level is concerned with cues and their immediate implications, such as hearing a gunshot close by and instinctively taking cover, the routine level deals with more sophisticated
expectations and implications of cues. For example, the series of actions one executes after making the decision to drive home—getting into one’s car, putting the key into the ignition, and turning the key with a foot on the brake.

As in the reactive level, individual differences can be analyzed at the routine level in terms of approach, avoidance and conflicts. Consider individuals at a party. In this case, a conceivable routinized behavior is the act of approaching and talking to a stranger. Extraverts, having a high sensitivity to cues for reward, may have learned that a stimulating conversation is rewarding, and so tend to engage in such approach behaviors. On the other hand, individuals who are highly sensitive to punishing cues tend to be afraid of approaching others at parties (Costa & McCrae, 1980) and so tend to engage in routine avoidance behaviors such as staying away from large groups. Individuals who are sensitive to conflicts (who can also, independently, differ in their sensitivities to cues for reward and punishment) easily feel frustration or annoyance if a conversation turns out to be less stimulating than expected, or feel anxious and unsure when the conversation partner shows signs of boredom, prompting the individual to try even harder to make the conversation work. This latter case should be differentiated from ones in which an individual is very sensitive to cues for punishment, in which case the individual will likely back off and try to avoid conversation altogether.

The Reflective Level

Often known also as the deliberative level, the reflective level is the home of high-level cognitive processes such as planning and conscious reasoning. The reflective level functions as the overall executive control that ‘oversees’ the operation of the lower levels. However, we want to suggest that the reflective level also embodies the same triadic AAC structure with individual differences in the corresponding sensitivities.

Of course, appeal to the approach-avoidance dyad in studies of motivation is anything but a new idea. It can be found in numerous theories (see Elliot 1999 for review), and it is widely recognized that individuals differ in their sensitivities to reward and punishment on the reflective level (Carver & White, 1994; Torrubia et al., 2001). Although most such studies have been designed to assess aspects of RST (which is reactive) in humans, the items in instruments used to do this in fact tap into routine and mainly reflective level processes. For example, the Sensitivity to Punishment, Sensitivity to Reward questionnaire (SPSR; Torrubia et al., 2001), includes items such as

- Does the good prospect of obtaining money motivate you strongly to do some things?
- Do you often renounce your rights when you know you can avoid a quarrel with a person or an organization?
- Do you think a lot before complaining in a restaurant if your meal is not well prepared?

The behaviors that correspond to such items obviously involve very reflective processes. Similarly, Regulatory Focus theory (Higgins, 1997) shows that there are chronic individual differences in motivation. A promotion-focused individual is concerned with nurturance-related motivations and is therefore sensitive to cues for reward. In contrast, a prevention-focused individual focuses more on security related needs, resulting in a bias toward cues for punishment. Regulatory Focus theory has been studied in a wide variety of contexts including goal pursuit and moral judgments, indicating that at least the approach-avoidance structure and individual differences in sensitivities to different classes of cues do exist on the reflective level.

However, the conflict system and individual differences in sensitivity to conflicts have received much less attention than the approach-avoidance pair, even though it is just as important an aspect of the motivation system. As already mentioned, conflicts arise within and between the approach and avoidance systems, but an individual’s sensitivity to conflicts is independent of the sensitivities to cues for reward and punishment. It might be tempting to equate sensitivity to conflict with sensitivity to cues for punishment, but as the sample item from the SPSR questionnaire about complaining in a restaurant indicates, there is a difference between experiencing the conflicting desires of seeking redress and avoiding potential unpleasantness with the restaurant staff, on the one hand, and simply having a high desire to avoid unpleasantness, on the other. The experience of conflict triggers rumination about the choice that the individual faces, whereas the desire to avoid the unpleasantness of a confrontation could have been avoided by the person just leaving when the service was bad. Functionally, sensitivity to conflicts is the threshold that determines when a conflict is experienced and produces a separate class of affective states and behaviors from those that result from simple approach or avoidance. The rumination, anxiety, and heart-wrenching despair that arise when one is forced to make choices are the products of internal conflicts and are not mere amalgams of behavior or affect produced in the approach-avoid systems. If one thinks of approach and avoidance tendencies as having different activation levels, then the sensitivity to conflicts is the threshold above which activation levels of incompatible motivations are experienced as internal conflict, while sensitivities to cues for reward and punishment influence how fast the respective activation levels change.

Connecting the Levels

The structure described in the previous sections provides a general framework for organizing personality parameters and for suggesting how these parameters might influence processes on the reactive, routine, and reflective levels. In appealing to the AAC structure we, of course, do not mean that, for example, a reflective level module should be split into three; we are certainly not proposing an approach-planning module, an avoidance-planning mod-
ule and a conflict-planning module, each associated with a distinct brain region. Rather, we are suggesting that there exist at least three broad classes of parameters which influence an individual’s selection of and access to different classes of strategies, or memories, or knowledge. We argue that broad personality dimensions arise from systematic differences in the set points of these parameters and should be modeled as such in cognitive architectures.

The patterning of a person’s sensitivities is consistent across the different levels of processing. So, for example, a person who is highly sensitive to cues for reward reacts and learns faster to cues for reward, engages more readily in habitual behaviors that he or she associates with reward, and values high level achievement goals. The consistency also implies that the relative relationships between the different sensitivities are preserved across levels. Therefore, a person who has a relatively higher sensitivity to cues for punishment than rewards will exhibit this difference in sensitivities across the three processing levels. The absolute magnitude of the sensitivities on each level can differ, but the relationships should remain consistent. Inconsistencies in the biases could explain behaviors that might be viewed as uncharacteristic of a person, as for example, when a person who typically values safety and security indulges in a spur-of-the-moment risky behavior such as reckless gambling.

Another consequence of the AAC structure is that similar systems on the three levels might be more tightly coupled than they are to others. That is, other things being equal, an activated approach system on the reactive level is more likely to cause responses in the approach systems on the routine and reflective levels, acting as a mechanism for the endogenous generation of related higher level goals and actions (e.g., Cacioppo et al., 1993). Consider the case of being the target of a scathing remark. The immediate reactive response might be to lash out and perhaps even to retaliate physically. However, because of the fear of reprisal and possible physical harm to oneself, the immediate response is suppressed. The reaction to the stimulus (insult) can trigger higher level processes, for example, to devise an elaborate plan for exacting revenge that acts to suppress the reactive level urge. The reverse also holds, where higher level goals and values, being more persistent, bias the perception of and sensitivity to different cues at the reactive and routine levels. An example is the cognitive bias that results from different task framing which influences actual task performance. Anxious students who want to do well, but are afraid of being seen as incompetent, perform better when the task is reframed to emphasize its difficulty (Born et al., 2002; Weiner & Schneider, 1971). In the case of the reframed task, the system that deals with wanting to avoid appearing incompetent is less activated because the task is perceived as being highly difficult anyway, and therefore reduces conflict with the approach system, resulting in lower state anxiety and allowing approach system behaviors (e.g., persisting in performing the task) to manifest themselves.

Implementing the AAC triad on the Reactive Level

Our prototype implementation of the AAC structure on the reactive level is inspired by RST and is combined with the Cues-Tendency-Action (CTA) re-parameterization of the Dynamics of Action model (Revelle, 1986; Atkinson & Birch, 1970). CTA models the dynamic interaction between cues and tendencies within and between the approach, avoidance and conflict systems. In particular, it models the interaction between conflicting tendencies and actions. It also includes the feedback mechanism for the interaction of consummatory actions with the behavioral tendencies. The hybrid RST-CTA architecture is shown in Figure 1.

The model is implemented on a set of virtual characters using the Twig animation system (Horswill, 2009). Screenshots from the simulation are shown in Figure 2. The focus of our simulation is the yellow child, who interacts with the red child, the ball, and the yellow adult–his parent. The yellow child perceives the other “objects” (i.e., the red child, ball, and adult) in the environment as stimuli. An input stimulus perceived by the agent (yellow child) is a tuple comprising the object, the object’s action and the object’s distance from the agent (i.e., the red child, the ball, and the adult). An example of an input is I = (name, action, distance). For example, I = ((red child), play, 2.3) indicates that the yellow child sees the red child playing 2.3 distance units away.

The expectation module (Figure 1a) uses the input I to form an expectation about what type of cue the stimulus is along four dimensions––reward (R+), non-reward (R-), punishment (P+) and non-punishment (P-). For instance, the red child is a cue for both reward (R+, playmate) and punishment (P+, aggressive child), the degree of which is scaled by his action (R+ is higher if the red child is being friendly) and distance (an aggressive child is less threatening if he is further away). The agent also uses the current stimulus (I) and information he has about its current action (A) to generate an expectation of what he should expect at the next moment. For example, a hostile red child is expected to approach the agent aggressively after issuing a threat (see Figure 2a) and the expected action will be flagged as highly punishing (P+), and if the agent runs back to the adult, he will expect to attain a certain amount of safety when he is close to the adult (P-).

The behavioral tendencies (Figure 1b) react to the input stimulus based on how the stimulus is evaluated along the four dimensions (R+, R-, P+, P-) with different sensitivities. A stimulus can cause changes to more than one behavioral tendency. In the case of the red child who is both a R+ and P+ to the yellow child, the presence of the red child activates the yellow child’s tendencies to both approach and avoid. The module also responds to consummatory actions taken by the agent at a rate defined by the sensitivity (S) (Figure 1b). For example, the act of playing with a ball is a consummatory action that reduces the tendency to continue playing.

The BIS module (Figure 1c) detects conflicts in behavioral tendencies and expectations, and responds by activating information-gathering behavior and inhibiting the conflicting
tendencies (Figure 1g). In Figure 2a, the yellow child wants to approach the red child but is also afraid that the red child might hurt him, causing the yellow child to inhibit his actions and look nervously at the red child. A child with high sensitivity to conflict (trait anxiety) will pause more often. On the other hand, if the yellow child has low sensitivity to conflict and high sensitivity to reward (extraversion), he will engage in “riskier” behavior and approach the red child with less hesitation. The BIS module also sends learning signals to adapt expectations and actions (Figure 1e and 1f). For example, if the red child threatens the yellow child, the yellow child will expect physical aggression from the red child to follow. An expectation violation occurs when the red child does not follow up his threat with physical violence. This causes the yellow child to eventually learn that a vocal threat from the red child is less of a cue for punishment. If the red child is friendly instead of aggressive, a yellow child who is highly sensitive to reward will react more strongly to the R+ component of the stimulus and learn more quickly that the red child is much more of a cue for reward (Figure 2c).

The action selection module (Figure 1d) receives the activation levels of the behavioral tendencies and activates the actions corresponding to the behavioral tendency with the maximum activation level. The appropriate commands associated with the selected action, A, is passed to the Twig animation system.

**Conclusion**

Personality and individual differences are coherent patterns of thoughts, feelings, desires and behaviors within individuals. Inspired by RST, we have proposed that personality can be organized and investigated in terms of approach, avoidance, and, importantly, conflict on the reactive, routine, and reflective levels, where an individual is consistent in his/her sensitivity to different types of cues on all three levels. We believe that this organization is useful for examining the influence of personality on other cognitive processes such as planning and moral deliberation. We also hope that the proposed structure might be informative for applications that require integrating personality into AI systems, such as in interactive games and drama, and various kinds of simulations of, for example, interpersonal interactions.

**References**


and diverse disorders. Current Directions in Psychological Science, 18, 195-199.


Figure 2: Screenshots from Twig showing some of the behaviors the yellow child can exhibit with different biases in his sensitivities to different environmental cues.