A computational cognitive-affective model of decision-making

Christopher L. Dancy (christopher.dancy@bucknell.edu)
Department of Computer Science, 335 Dana Bldg
Lewisburg, PA 17837

David M. Schwartz (dms061@bucknell.edu)
Department of Computer Science, 335 Dana Bldg
Lewisburg, PA 17837

Abstract

How do affective processes interact with cognitive processes to modulate our behavior? Understanding the processes that influence the interactions between affective stimuli and human decision-making behavior is important for predicting typical behavior under a variety of circumstances, from purchasing behavior to deciding when to enact certain rules of engagement in battle scenarios. Though some computational process models have been proposed in the past, they typically focus on higher-level phenomena and are less focused on the particular architectural mechanisms related to the behavior explored. This, in turn, can make it very difficult to combine the proposed model with existing related work (i.e., the models can't be tractably combined).

We used a modified version of the Iowa Gambling Task to explore the effects of subliminal affective (visual) stimuli on decision-making behavior. We developed a model that runs within the ACT-R/Φ architecture that completes the same task completed by participants. In addition to the affective and cognitive memory components particularly important to the discussion, the model also uses perceptual and motor components within the architecture to complete the task. The architecture has representations of *primitive affect* that interact with cognitive memory components mainly through an *affective-associations* module (meant to capture behavior typically ascribed to several amygdalar substructures). The model and affective architectural mechanisms provide a process-oriented explanation for the ways affect may interact with higher-level cognition to mediate human behavior during daily-life.

Keywords: Cognitive Architecture; HumMod; ACT-R; Affect: IGT; Decision-Making; Emotion

Introduction

How do affective processes interact with cognitive processes to modulate our behavior? Though this question is important, we've only seen a relatively recent surge in computational process models that have explored this question (e.g., Marinier III et al., 2009; Marsella et al., 2010). Indeed, even Newell did not have *emotion* (and motivation) as a topic that was most important to address when developing a unified theory of cognition. As more evidence of the importance of emotional/affective processes has accumulated through experimentation and simulation, it has become clear that affect and emotion play a fairly central role in mediating behavior (e.g., Bechara et al., 1997; LeDoux, 2012; Panksepp & Biven, 2012).

We conceive of emotion as an interaction between *affective* and *cognitive* processes. When we make these distinctions we do so with the idea that the two categories describe both qualitative and quantitative differences in computational processes that, nonetheless, interact within a whole computational behavioral system (e.g., see Figure 1 that describes differences in *levels,* Panksepp et al., 2011). We see affective processes as those modulate subsymbolic representations within the cognitive system, which results in certain behavior that may be deemed as *emotional*.

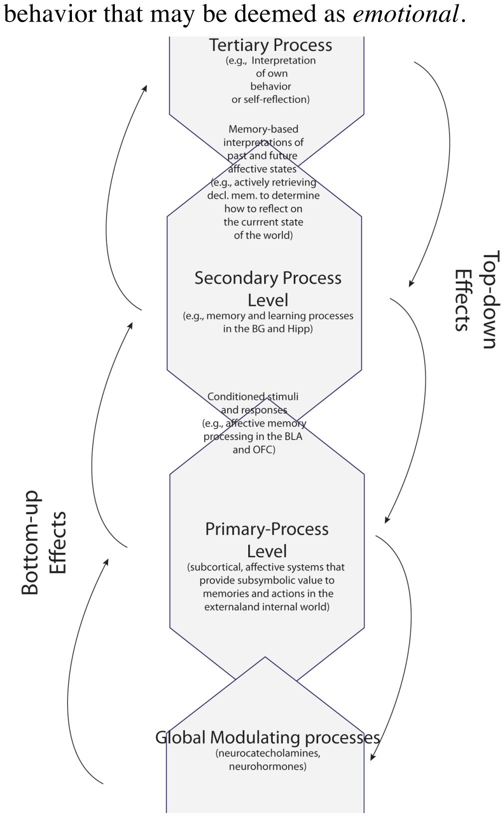

Figure 1. Levels of behavioral processes from Panksepp et al. (2011)

While some affective processes may have less quantitative effect on symbolic and subsymbolic representations depending on context, the implication here is that no such human interaction is truly without some bias due to affective processes. Ultimately other portions of a conditional context may have more effect on resulting action (e.g., a current *goal*/*intentional*), however these affective processes still may have small effects on the representations/actions that occur in a computational cognitive system. Put in a more high-level example, just because one may ultimately elect to buy the more economically functional vehicle, doesn't mean that the affective drive to select the new sports car did not factor into the decision.

Below, we give a description of a decision-making study and some results from this study. The study used a modified version of the Iowa Gambling Task (IGT) that involved subliminally presented visual stimuli to explore the particularly non-conscious and subsymbolic effects of affective processes. We also detail and discuss an affective-cognitive model of this task that uses components of the ACT-R/Φ architecture to represent interactions between affective and cognitive processes.

Description of IGT Study and Results

97 undergraduate students were recruited as participants for this study (52 males and 45 females). The average ages of males and females were similar at 20.7 and 19.8 (respectively). Electrodermal Activity (EDA) data were collected for the final 66 (37 males and 29 females) participants (data not reported here). All participants were given college course extra credit for participation.

A filter process that removed participants who completed less than 20% of their trials due to time restrictions (max 3.5s per trial) resulted in the removal of 4 participants' data from further analysis; data from 93 total participants were analyzed. The *negative*, *neutral*, and *positive* (image) groups each had 31 participants. We ceased participant enrollment in the study after we crossed a 31 per-group threshold for task-related behavioral analysis and all volunteers had the opportunity to participate.

Participants used a version of the IGT that included a fixed reward and punishment schedule for each deck that was the same as the schedule used for the original IGT by Bechara et al. (2000). A modified computerized version of the IGT was used that runs in Matlab and uses the Psychtoolbox Matlab extensions (Brainard, 1997). Psychtoolbox extensions were used due to their high timing accuracy, community support, and cross-platform availability and the specific software used has had IGT-specific timing tests done to confirm timing accuracy (Dancy & Ritter, 2016).

The visual stimuli presented during the IGT were obtained from the International Affective Picture System (IAPS; Lang et al., 1997). Table 1 lists the images used in image sets used by the different groups. Male and female pictures were matched so that, for each group, they had similar valence/arousal/dominance ratings and had a similar content subject; for example, some snake pictures had different ratings between sexes within the IAPS manual, so those images with lower valence/higher arousal ratings among the same category were chosen. Given that picture ratings in all categories differed between sexes, this method allowed more consistency in mean measured quantitative ratings among participant sexes.

Table 1. The IAPS images (and the accompanying average valence, arousal, and dominance rating) used in the experiment.

Picture-Set	Picture Numbers
$\text{Negative}_{\text{Male}}$	1050, 1202, 1220, 1304, 1525
$\text{Negative}_{\text{Female}}$	1050, 1120, 1201, 1202, 1525
$\text{Neutral}_{\text{Male}}$	1670, 7006, 7010, 7080, 7175
$\text{Neutral}_{\text{Female}}$	1670, 7004, 7010, 7012, 7175
$\text{Positive}_{\text{Male}}$	4180, 4210, 4232, 4664, 8501
$\text{Positive}_{\text{Female}}$	4505, 4525, 4660, 8001, 8501

Before participating in the study, all participants read and signed a consent form approved by the Office of Research Protections (ORP) at Penn State. Participants were assigned to one of three possible groups (with different accompanying treatments): a *negative group* with a negative image treatment, a *neutral group* with a neutral image treatment, or a *positive group* with positive image treatment. Images (consistent with participant sex and group) were presented to participants for 17ms after deck selection if they selected from one of the *bad decks* (those that give a negative net amount of money) and plain gray images were presented for the same amount of time if a card selection was made from one of the other two decks. For a full explanation of the typical IGT procedure, see (Bechara et al., 2000).

Results

As with previous IGT-based studies we split deck selection analysis into five blocks, 20 deck selections per block. *Score* was calculated by subtracting the total number of card selections from decks A and B (the *bad decks*) from the total number of card selections from decks C and D.

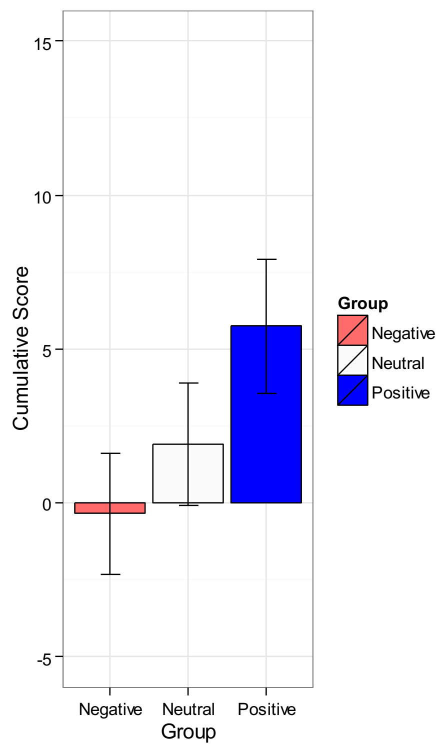

Figure 2. Cumulative score (±SEM) for all participants after the final block

Participants in the positive image group showed the highest score (Figure 2) when averaged across blocks, but all groups had a positive score by the final block (Table 2). Scores increased for all groups from blocks 1-3 and blocks 4-5, but decreased from blocks 3-4.

Table 2. Mean score of participants in all of the blocks by group. Standard errors are in presented in the parenthesis

Group	B1	B2	B3	B4	B5
Negative	-4.2	0.2	1.1	-0.2	2.7
	(1.4)	(1.3)	(1.5)	(1.3)	(1.4)
Neutral	-3.5	-0.3	2.4	1.5	1.8
	(1.1)	(0.9)	(0.8)	(1.0)	(1.1)
Positive	-3.3	-0.5	3.0	2.6	4.0
	(1.3)	(1.2)	(1.2)	(1.1)	(1.0)

A $3X5$ (group by block) mixed factor ANOVA of participant score revealed a highly significant effect of block ($F(4,\ 360)\ =\ 13.22, p\ <\ .0001$) on score, however it did not reveal a significant group ($F(2,\ 90)\ =\ 0.81,\ p\ =\ .4$) or a group:block interaction ($F(8,\ 360)\ =\ 0.40,\ p\ =\ .9$) effect.

When sex is also taken into account, males and females show an opposite score distribution across groups (Figure 3).

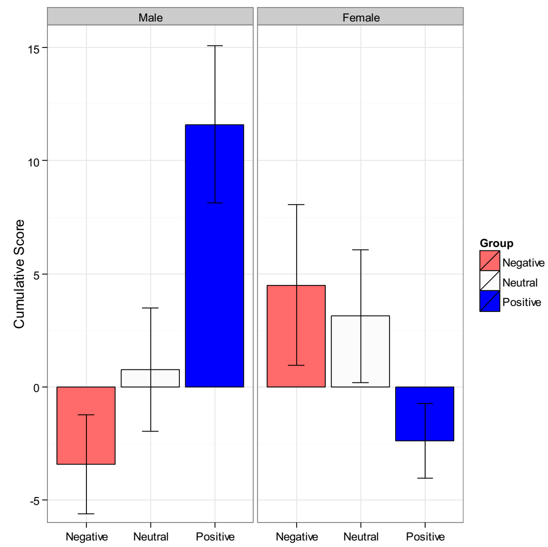

Figure 3. Cumulative score for male (left) and female (right) participants after the final block.

Among male participants, those in the positive group showed the highest cumulative score and those in the negative group showed the lowest score (the only negative among male participants). Conversely, among female participants, those in the positive group received the lowest score (the only negative among female participants), while those in the negative group received the highest scores.

The decision-making model

To simulate this task and potentially understand more about the processes that mediates behavior during this task (and others that show some effects of subliminal affective stimuli), we developed a cognitive-affective model that runs within the ACT-R/Φ architecture. This model uses simulated eyes and hands to perceive the task (e.g., see the decks, cards, rewards, and affective images) and provide feedback (e.g., press a key to select a card from a deck). To make decisions, the model uses both procedural and declarative memory (Figure 4).

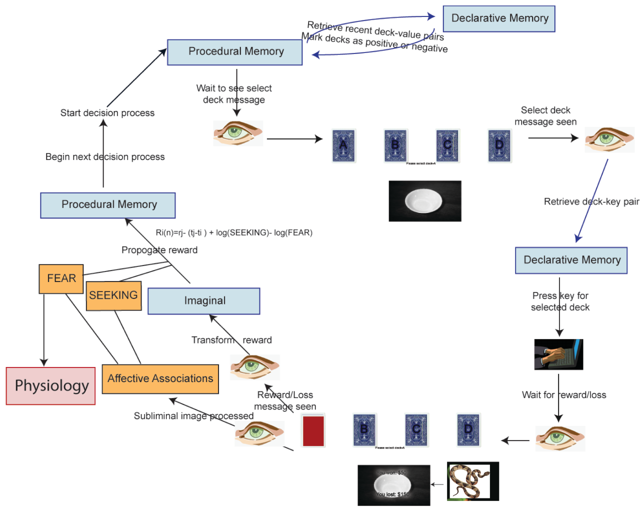

Figure 4. A high-level diagram of the ACT-R/Φ model

After the model has made a deck selection and a reward/loss is shown, it uses those values to reinforce the utility of those production rules recently fired:

$$U_i(n) = U_i(n-1) + \alpha\big(R_i(n) - U_i(n-1)\big) \quad [1]$$

$$R_i(n) = r_j - \left(t_j - t_i\right) + \log(Value_{SEEKING}) - \log\,(Value_{FEAR}) \quad [2]$$

$$val(t) = W_{rew} * rew(t)^{\gamma} \text{ - } W_{aversion} * loss(t)^{\gamma} \quad [3]$$

Here, $U_i(n-1)$ represents the current utility value, α is a learning rate, and $R_i(n)$ is a reward that is determined by equation 2. In equation 2 r_j represents the reward received and $t_j - t_i$ is the temporal discount that is given to a reward so that the length of time between reward onset and a rule firing determines how much reward is applied to the utility value. $Value_{SEEKING}$ and $Value_{FEAR}$ in equation 2 are reward offsets that take into account the current affective state of the model (see Dancy, 2013) for some description of the modules/systems in ACT-R/Φ that control these values. Though the SEEKING system can be affected by several things in a realistic environment (e.g., the model would see an increase in SEEKING activation/value if it were thirsty), the limited scope of this model means that the SEEKING and FEAR values are practically determined by the emotional images flashed after selecting a card from a *bad deck*; this is controlled by the affective-associations module in ACT-R/ Φ (which is shown in Figure 4. The affective value for the images flashed is derived from equations 4 and 5, which use the values for arousal, valence, and dominance from the images listed in Table 1 (specific values are available in the IAPS manual, Lang et al., 1997).

$$FEAR_{value} = \frac{(arousal + \varepsilon) * (10 - (valence + \varepsilon) - (dominance + \varepsilon))}{90} \quad [4]$$

$$SEEKING_{value} = \frac{(arousal + \varepsilon) * ((valence + \varepsilon) + (dominance + \varepsilon) - 10)}{90} \quad [5]$$

The actual reward (i.e., r_j in equation 2) is determined by a function that transforms the gain and loss that results from selecting a card from a deck on a given trial, using the imaginal/imaginal-action buffers. This function implements equation 6 below, which is a slightly modified version of an equation discussed by Ahn et al. (2008) and proposed by Napoli and Fum (2010) to be used in ACT-R.

$$val(t) = W_{rew} * rew(t)^{\gamma} \text{ - } W_{aversion} * loss(t)^{\gamma} \quad [6]$$

These equations all reinforce production rules, which can cause some production rules to be less likely to fire over time (those that consistently have a negative reward, and consequently a lower utility will decrease in likelihood of firing overtime). This selection rule is encapsulated in equation 7.

$$P(i) = \frac{e^{U_i/\sqrt{2}s}}{\sum_j e^{U_j/\sqrt{2}s}} \quad [7]$$

$P(i)$ is the probability of selecting rule i which is determined by comparative weight of the rule i as well as any *procedural noise* (represented as *:egs* in canonical ACT-R).

The model also uses declarative memory to learn and make decisions. It encodes deck-value pairs and these pairs ultimately control which decks are selected. At the beginning of making a deck selection, the model queries declarative memory for deck-value pairs for each of the decks. The declarative memory elements with the highest activation (governed by equation 8) are selected.

$$A_i = B_i + S_i + P_i + \varepsilon_i \quad [8]$$

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) + \beta_i \quad [9]$$

Thus, the model uses the majority of the major components of the original ACT-R architecture (including perceptual and motor systems).

Model Comparison to Study Results

The IGT model was run a total of 360 times, 120 for each of the negative, neutral, and positive image groups. Half of the model runs within each group were male, while the other half were female. Thus, this resulted in 60 unique runs of the model. Because the model was originally developed as a prediction of the processes occurring and those behavioral data that result from such processes, sex-based differences were not a focus. In the model, the distinction between male and female only comes into play with the affectively valued visual stimuli (which nonetheless have very similar values.)

For all groups the, the model predicted a similar scoring trend (positive) from blocks 1 to 3 (Figure 5). Overall the negative model seemed to deviate the most from the actual observed score by participants across decks.

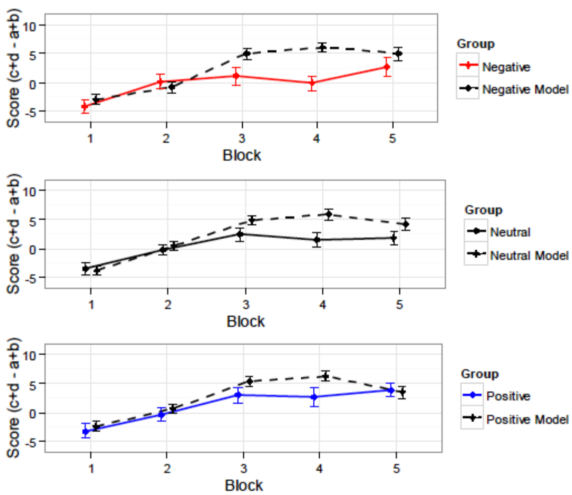

Figure 5. Comparison of score performance between participants and the model for negative, neutral, and positive groups

The neutral model predicted the scores for the five blocks best followed by the positive and negative models (Table 3).

Table 3. Comparison between model predictions for the different groups.

Model/Group	r^2	*RMSD*
$IGT_{Negative}$/Negative	.56	3.48
$IGT_{Neutral}$/Neutral	.94	2.49
$IGT_{positive}$/Positive	.81	2.05
All	0.72	2.74

Discussion and Conclusion

The model fit best to those data from the participants in the neutral group best, though the model did also fit reasonably well to those data from the positive group. It would seem that there is a key point of change that the model does not exhibit (i.e., in block 4). The model continues on the positive trend, as exhibited in previous blocks, while participants show a dip in performance during this block. Because the model does not switch deck selection in the same way participants do (and thus, continues on a *greedy* path), the model tended to exhaust decks at a certain point, causing the dip in performance seen in the final block.

The model appears to have underestimated the effects of the subliminally presented affective stimuli. While, the affective stimuli did have certain subtle (subsymbolic) effects through the affective-associations module, those participant data showed a much more overt effect on performance. What's more, these behavioral effects seemed to have some dependency on participant sex, for which the model had very little account.

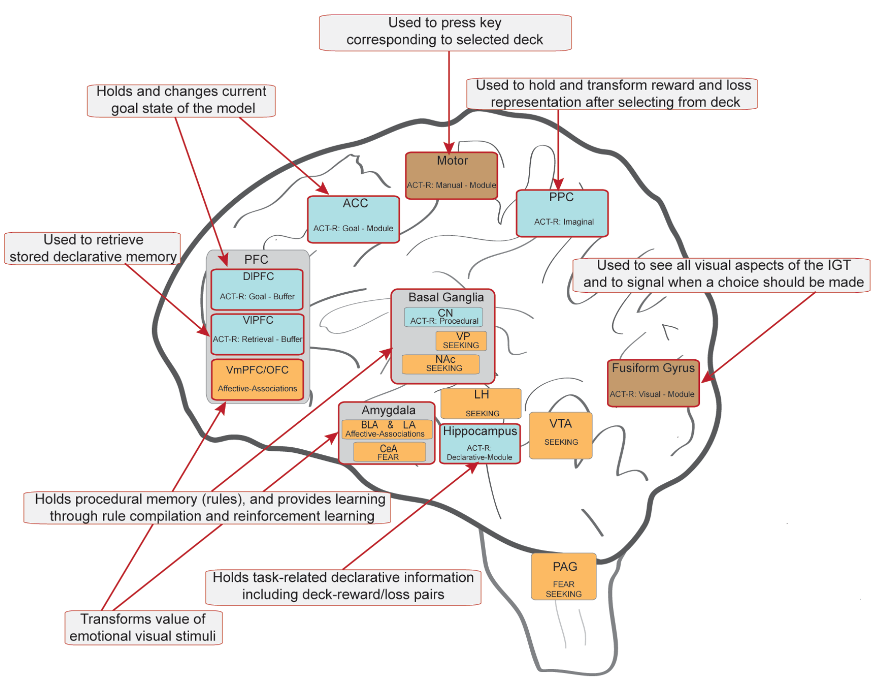

Figure 6. Predicted brain areas and main functions by the model/architecture.

Though the model did not predict several aspects of these presented data, it provides a useful framework for future work and related simulation. Indeed, using a system like ACT-R/Phi for the simulation also allows one to provide early predictions of brain areas involved in related affective decisions (Figure 1). This is due to the architectures use of ACT-R theory (which has various functional modules that have been associated with certain neural structures, Anderson, 2007) and theory from affective neuroscience (e.g., Panksepp, 1998; Panksepp & Biven, 2012). The predictions from Figure 6 can be further explored in future studies. Future plans for this particular model include running a ranging parameter sweep on potentially varying parameters (e.g., W_{rew} and $W_{aversion}$) to see if the model can more closely fit to these data presented here.

Existing theory, data, and these data presented here make it clear that affective processes can have an overt effect on decision-making behavior, even when the affective stimuli causing the activation of such processes isn't overt. It is important to understand these effects as they can be useful to positively, or negatively, influence our decisions in various ways that may fail to reach our awareness. The model and mechanisms presented provides a first step towards providing a more systematic and unified account of the modulating effects of affective stimuli on cognitive behavior.

Acknowledgments

Some of this work was supported by the Sloan graduate scholarship and the Bucknell Scholarly Development Grant.

References

Ahn, W., Busemeyer, J. R., Wagenmakers, E., & Stout, J. C. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science, 32*(8), 1376-1402.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: OUP.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex, 10*(3), 295.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*(5304), 1293-1295.

Brainard, D. H. (1997). The Psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Dancy, C. L. (2013). ACT-RΦ: A cognitive architecture with physiology and affect. *Biologically Inspired Cognitive Architectures, 6*(1), 40–45.

Dancy, C. L., & Ritter, F. E. (2016). IGT-Open: An open-source, computerized version of the Iowa Gambling Task. *Behavior research methods*, 1-7.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International Affective Picture System (IAPS): Technical manual and affective ratings*. The Center for Research in Psychophysiology, University of Florida. Gainesville, FL.

LeDoux, J. E. (2012). Rethinking the Emotional Brain. *Neuron, 73*(4), 653-676.

Marinier III, R. P., Laird, J. E., & Lewis, R. L. (2009). A computational unification of cognitive behavior and emotion. *Cognitive Systems Research, 10*(1), 48-69.

Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. In K. R. Scherer, T. Banziger, & E. B. Roesch (Eds.), *A blueprint for affective computing: Crossfertilization between emotion psychology, affective neuroscience, and affective computing,* . New York, NY: OUP.

Napoli, A., & Fum, D. (2010). Rewards and punishments in iterated decision making: An explanation for the frequency of the contingent event effect. *In proceedings of the International Conference on Cognitive Modeling*, Philadelphia, PA.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York, NY: OUP.

Panksepp, J., & Biven, L. (2012). *The Archeology of Mind: Neuroevoloutionary Origins of Human Emotions*. New York, NY: W.W. Norton & Company.

Panksepp, J., Fuchs, T., & Iacobucci, P. (2011). The basic neuroscience of emotional experiences in mammals: The case of subcortical FEAR circuitry and implications for clinical anxiety. *Applied Animal Behaviour Science, 129*(1), 1-17.