Cogn Process (2011) 12:375–378
DOI 10.1007/s10339-011-0401-4

LETTER TO THE EDITOR

Could positive affect help engineer robot control systems?

Markus Quirin · Joachim Hertzberg ·
Julius Kuhl · Achim Stephan

Received: 22 September 2010 / Accepted: 21 March 2011 / Published online: 8 April 2011

Abstract Emotions have long been seen as counteracting rational thought, but over the last decades, they have been viewed as adaptive processes to optimize human (but also animal) behaviour. In particular, positive affect appears to be a functional aspect of emotions closely related to that. We argue that positive affect as understood in Kuhl's PSI model of the human cognitive architecture appears to have an interpretation in state-of-the-art hybrid robot control architectures, which might help tackle some open questions in the field.

Recent years have witnessed attempts to implement emotions or at least some functional (rather than experiential) aspects of emotions in robots (e.g., Breazeal and Brooks 2005; Pfeifer and Bongard 2007; Trappl et al. 2002). By now, this endeavour has predominantly focused on how to implement the ability to encode and/or decode affective signals to facilitate interactions with humans. Lately, however, attempts have been made to adopt another important function of emotions to robotics, namely to organize behaviour by modulating attention, selection and learning (Arbib and Fellous 2004). In this vein, based on a functional approach towards affect-cognition interactions considering brain functionality, we put forth here the personality systems interactions (PSI) model (Kuhl 2000) that describes how changes in positive affect modulate the interaction between planning and executive systems to optimize human behaviour. We propose that, and later argue why, the PSI model might also be applied to robots.

Many cognitive architectures incorporate both a *planning system* and an *executive system* (see Fig. 1). The planning system develops plans for the attainment of specific goals on the basis of symbolic propositions and deliberate, deductive reasoning. It stores these plans in the form of intentions until situations are encountered that are appropriate for their enactment. Typically, it is activated only when target-aimed behavioural routines are not immediately accessible. By contrast, the *executive system* prepares the initiation of intended actions, monitors their execution and supports error detection (and to some extent error correction as long as it can be done without interrupting ongoing behaviour and reverting to the planning system). Typically drawing on parallel-distributed processing, it organizes a great amount of contextual information for fast and context-sensitive online control of actions on the basis of a preparatory schema of the *feedback* expected as a result of the intended action (Holmes et al. 2004, for neuroscientific evidence). Such a type of control is much more efficient and less time-consuming than comparing feedback from already executed behaviour with a feedforward control signal that would involve two incompatible codes (Hommel et al. 2001). Many

M. Quirin (✉)
Institute of Psychology, University of Osnabrück,
Seminarstraße 20, 49069 Osnabrück, Germany
e-mail: mquirin@uos.de

J. Hertzberg
Institute of Computer Science, University of Osnabrück,
Albrechtstr. 28, 49076 Osnabrück, Germany
e-mail: joachim.hertzberg@uos.de

J. Kuhl
Institute of Psychology, University of Osnabrück,
Seminarstraße 20, 49069 Osnabrück, Germany
e-mail: jkuhl@uos.de

A. Stephan
Institute of Cognitive Science, University of Osnabrück,
Albrechtstr. 28, 49076 Osnabrück, Germany
e-mail: acstepha@uos.de

corrections can thus be done online without necessitating behavioural interruption and returning control to the planning system.

What could be the role of emotions in this context? On the basis of the PSI model of personality functioning, we propose that emotional states can modulate the decision as to whether error correction can be done without interrupting ongoing behaviour or whether control has to be reverted to the planning system. The heuristic value of this notion may become especially apparent when considering an intriguing problem that waits for future solutions in both cognitive psychology and robotics, namely the control flow between planning and executive systems.

The PSI model was originally developed to solve putatively paradoxical findings of experimental studies on the enactment of plans in humans. Specifically, it has been found that humans typically have a memory advantage for uncompleted over completed tasks (Zeigarnik 1927), i.e., they tend to remember not yet accomplished tasks but tend to forget already accomplished tasks. Paradoxically, it has been found that memory for uncompleted tasks is better for depressive individuals (Johnson et al. 1983) or those with depression-like symptoms such as procrastinators (Goschke and Kuhl 1993). Why do individuals typically being most inefficient in enacting plans show superior memory for them? The PSI model postulates that the answer may lie in an impaired interaction between the planning and executive systems rather than in the functionality of the planning system itself (as might be indicated by impaired memory). More so, the PSI model postulates that planning and executive systems have a default inhibitory connection even in healthy individuals: The more the agent engages in planning, the lower is the current activity and the more inert (and thus effortful) is the reactivation of the executive system. Vice versa, the more the agent engages in online control of task performance, the lower is the activity (and the more inert the reactivation) of the planning system (dotted arrow in Fig. 1).

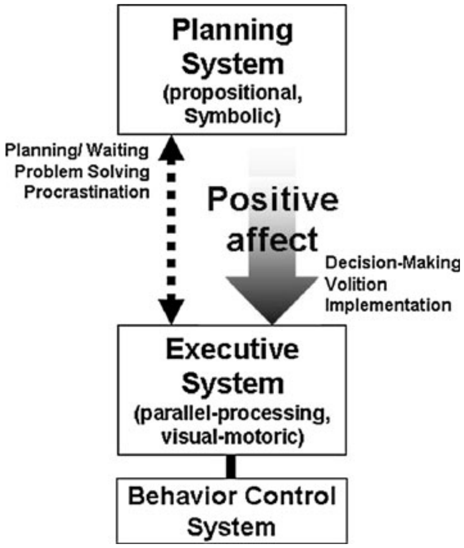

Fig. 1 Three-layer cognitive architecture as inspired by Kuhl's PSI model of human cognition. *Note.* As a default, a mutual inhibition (*dotted arrow*) between planning and executive system is hypothesized that can be disinhibited by PA in a way that the planning system obtains more control over the executive system (*descending arrow*). The focus is on the relationships between the planning and the executive system

Based on the fact that depressed individuals typically lack positive affect states such as joy or power, the PSI model postulates by implication that in healthy individuals the inertia of the planning system's activity can be counteracted by an *increase* of positive affect to the effect that plans can be enacted earlier (Kuhl and Kazén 1999, for empirical evidence). This model also accounts for the finding that procrastinators and depressives often miss out to enact a plan even when it is completed and the situational conditions are optimal to enact it (Goschke and Kuhl 1993). Additional symptoms of procrastinators or depressives relate to other properties of the executive system, such as problems with initiating the *first step* of an action sequence (executing the remaining steps is easier once the first step is done), problems with context-sensitive error detection and correction (to the extent that the same error is repeated even if the actor realizes what is wrong), problems with context-sensitive decisions concerning exceptions (e. g., that a deviation from the plan is indicated).

The notion that positive affect increases the interaction between planning systems and executive systems is also supported by neuroscientific evidence. Specifically, the left prefrontal cortex (typically supporting symbolic planning) and right parietal cortex (typically processing sensory-motor feedback of information about spatial parameters of objects and how to grasp them) show reduced coherence in electroencephalographic alpha frequency power for depressed individuals, which can be increased by testosterone, a hormone linked to power and positive affect (Schutter et al. 2005). In sum, based on theoretical notions and empirical evidence, the PSI model assumes that adaptive plan execution in humans can be modelled on the basis of an inhibitory connection between a planning and an executive system that can be disinhibited by positive affect such that flow of plan-related information from the first to the latter system is facilitated.

Here, an interesting question arises as to whether this type of functionality might be useful for artificial agents or robots. Having an executive layer in robot control is state of the art in hybrid robot control systems (Kortenkamp and Simmons 2008, for overview), where the executive is interfacing a symbolic planning layer and the behavioural control layer that is responsible for generating physical action. Similar to human mind architectures such as the PSI model, the role of the executive is to translate flexibly the relatively high-level, abstract advice that symbol-based action plans would provide, into control routines running in closed loop on robot hardware, taking the current situation into account, monitoring execution and handling exceptions (Kortenkamp and Simmons 2008, p. 197). The above-mentioned problems of people who cannot easily generate the positive affect that is necessary in humans to make the transition from planning to execution are indeed reminiscent of a conceptual problem known in hybrid robot control architectures: specifying the control flow between the planning and the executive layers.

For example, handling exceptions from a preset plan, which includes detecting these exceptions in the first place, is a major problem in robotics. If the number and precise set of categories of possible exceptions in executing a high-level planned action is assumed to be known in advance, then the executive system can be implemented as some variant of finite state automaton—which is in fact done with success in several working hybrid robot control systems (Kortenkamp and Simmons 2008, ch. 8.3.3). In all other cases, the current state of the art in robotics has in fact no good answer to the question of how to build an executive system in some principled way and how to interface it to the planning layer on the one side and to the behavioural control on the other side. Difficult questions to answer within the control system include: What precisely is an exception, given some particular abstract action to be executed in a dynamic environment, and given the recent stream of raw sensor data? If an exception occurs, when should control resume from the behavioural control to the executive? Once a physical behaviour is determined to have failed, when should it be retried, and when should an alternative be sought? When should this alternative be generated relatively locally by the executive itself, and when should control be given back to the planning layer to re-plan? Although a number of successfully deployed robot systems exist where these problems have been solved heuristically (Kortenkamp and Simmons 2008, for overview), there is currently no solid theory for answering them in a principled way.

Without providing a definite solution to these problems, we propose that advances towards solutions may be made by considering the functional role of positive affect in humans as, for example, proposed by the PSI model. Specifically, humans often use a fuzzy, holistic heuristic for assessing plan-state deviations or decisions about whether to stay with or to abandon a plan (Gros 2010; Ziemke and Lowe 2009). Following the PSI model, changes in positive affect may function as an indicator of this assessment by collecting all relevant information about the amount of resources that should be raised for re-planning (e.g., when severe difficulties arise) as compared to the expected costs and risks involved in retrying by simply applying the online error correction capacity of a well-developed executive system. The amount of change in positive affect does not only depend on the deviation between anticipated and obtained results, but also on the amount of internal resources available to deal with the deviation (and perhaps on the type of error and contextual information that help assess the chances of online correction versus re-planning). Thus, the level of positive affect ‘informs’ about the net result of a variety of conditions that help assess the chances and risks of continuing online control versus re-planning (Schwarz and Clore 1983). In robotics, a functional analogue of positive affect would of course need another type of implementation than in humans. This would amount to a holistic management of internal resources which is indeed a relevant matter in robots. For example, when an internal resource like battery power is high, much can be invested in the power-consuming motor system for trial-and-error efforts, whereas control should favour power-efficient planning efforts when power is low.

The point here is whether a particular control decision within a hybrid robot control system would be dominated or triggered by reaction or deliberation, grossly depends on the granularity and time scale of what this decision is about (Kortenkamp and Simmons 2008, ch. 8.2.2). For example: ‘Don’t even start planning if you need to break *now*; don’t even try to optimize your daily errands on the basis of millisecond cycle time slots.’ These are simple cases that can nicely be handled in state-of-the-art architectures. However, a holistic parameter as an analogue to positive affect, which mirrors a variety of internal and external state features as far as sensed by the robot, may help make decisions about passing control between layers in exception or failure situations as mentioned above. These decisions are notoriously hard to make from the local view of a single control layer (e.g., reactive, sequencing, deliberative layers), so positive affect might induce an additional, independent perspective.

In sum, in robot control architectures, analogues of positive affect might flexibly modulate the switching of control between executive and planning layers based on an integrated assessment of amount of internal resources and perceived current fit of the planner’s symbolic world model and state of affairs during behaviour execution—thereby addressing one of the open questions in robot control architectures.

Despite these tentative suggestions as to the functional role of affect at the plan-executive interface, the question regarding the functionality of emotions in robots remains wide open. When reflecting about this question, it should be kept in mind that the PSI model, although basically being a psychological model, already focuses on the functional rather than the experiential aspect of affect. A major advantage of this approach is that artificial implementations of brain or mind-based operations can be detached from the physical medium of the prototype, which might not be the case if experiential aspects are taken into account. We plan to elaborate the potential benefits of the experiential aspect of affects for artificial architectures in future work.

References

Arbib MA, Fellous J-M (2004) Emotions: from brain to robot. Trend Cogn Sci 8(12):554–561

Breazeal C, Brooks R (2005) Robot emotion: a functional perspective. In: Fellous J-M, Arbib MA (eds) Who needs emotions? The brain meets the robot. Oxford University Press, New York, pp 271–310

Goschke T, Kuhl J (1993) Representation of intentions: persisting activation in memory. J Exp Psychol Learn Memory Cogn 19(5):1211–1226

Gros C (2010) Cognition and emotion: perspectives of a closing gap. Cogn Comput 2:78–85

Holmes NP, Spence C, Giard M-H, Wallace M (2004) The body schema and multisensory representation(s) of peripersonal space. Cogn Process 5(2):94–105

Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The theory of event coding (TEC): a framework for perception and action planning. Behav Brain Sci 24(5):849–937

Johnson JE, Petzel TP, Hartney LM, Morgan RA (1983) Recall of importance ratings of completed and uncompleted tasks as a function of depression. Cogn Therapy Res 7:51–56

Kortenkamp D, Simmons R (2008) Robotic systems architectures and programming. In: Siciliano B, Khatib O (eds) Handbook of robotics, chap 8. Springer, Berlin, pp 187–206

Kuhl J (2000) A functional-design approach to motivation and self-regulation: the dynamics of personality systems and interactions. In: Boekaerts M, Pintrich PR (eds) Handbook of self-regulation. Academic Press, San Diego, CA, pp 111–169

Kuhl J, Kazén M (1999) Volitional facilitation of difficult intentions: joint activation of intention memory and positive affect removes strop interference. J Exp Psychol General 128(3):382–399

Pfeifer R, Bongard J (2007) How the body shapes the way we think: a new view of intelligence. MIT Press, Cambridge, MA

Schutter DJLG, Peper JS, Koppeschaar HPF, Kahn RS, van Honk J (2005) Administration of testosterone increases functional connectivity in a cortico-cortical depression circuit. J Neuropsychiat Clin Neurosci 17(3):372–377

Schwarz N, Clore GL (1983) Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. J Personal Social Psychol 45:513–523

Trappl R, Petta P, Payr S (2002) Emotions in humans and artifacts. MIT Press, Cambridge, MA

Zeigarnik B (1927) Über das Behalten erledigter und unerledigter Handlungen. Psychologische Forschung 9:1–85

Ziemke T, Lowe R (2009) On the role of emotion in embodied cognitive architectures: from organisms to robots. Cognitive Comput 1:104–117