Review

CellPress

Inferring affect from fMRI data

Brian Knutson, Kiefer Katovich, and Gaurav Suri

Department of Psychology, Stanford University, Stanford, CA 94305, USA

Neuroimaging findings are often interpreted in terms of affective experience, but researchers disagree about the advisability or even possibility of such inferences, and few frameworks explicitly link these levels of analysis. Here, we suggest that the spatial and temporal resolution of functional magnetic resonance imaging (fMRI) data could support inferences about affective states. Specifically, we propose that fMRI nucleus accumbens (NAcc) activity is associated with positive arousal, whereas a combination of anterior insula activity and NAcc activity is associated with negative arousal. This framework implies quantifiable and testable inferences about affect from fMRI data, which may ultimately inform predictions about approach and avoidance behavior. We consider potential limits on neurally inferred affect before highlighting theoretical and practical benefits.

Background and definitions

Although Galileo Galilei did not invent the telescope, he did refine its resolution enough to visualize the orbiting moons of Jupiter, allowing him to overthrow the then dominant geocentric view of the universe. Galileo's incremental innovation illustrates that inventing new measures is not sufficient to promote scientific advance; the resolution of those measures must also match the scale of the phenomenon of interest.

More recently, researchers have developed methods with enhanced resolution for peering not only into outer space, but also into inner space. Given that mental states can change rapidly, capturing their traces requires temporal as well as spatial precision. Different neuroimaging techniques have historically presented varying tradeoffs between temporal and spatial resolution. Whereas electroencephalography (EEG) provides good temporal resolution (e.g., approximately milliseconds), electrodes outside the skull have limited spatial resolution, particularly for signals that emanate from below the cortex. Conversely, whereas positron emission tomography (PET) affords subcortical spatial resolution (e.g., approximately millimeters), its temporal resolution is limited (e.g., approximately minutes or more). Near the end of the 20th century, fMRI grew in popularity, partially because it offered both spatial resolution (e.g., approximately millimeters, even for subcortical regions) and temporal resolution (e.g., approximately seconds). The spatiotemporal resolution of fMRI may better match that of neural activity associated with some psychological phenomena, including affective experience.

Along with the rising popularity of fMRI, scientific interest in the neural basis of affect also increased. Since Wilhelm Wundt's prescient writing over a century ago, 'affect' has referred to a broad range of phenomena including mood, emotion, and motivation (differing in their duration, causes, and consequences) [1]. Over 100 years of psychometric analyses of self-reported emotional experience have vindicated Wundt's claim that affect varies along at least two independent dimensions: valence (running from good to bad) and arousal (running from high to low) [2]. Whereas statistical analyses alone cannot specify which rotation of these dimensions best describes affective space [3], later theorists advocated a 45° rotation of the valence and arousal axes, transforming them into dimensions of 'positive arousal' (running from feelings of excitement to boredom) and 'negative arousal' (running from feelings of anxiety to calm) [4]. Indeed, most sensory stimuli (e.g., olfactory, gustatory, auditory, or visual) invoke a 'V'-like pattern of affective reactions that falls along these positive arousal and negative arousal dimensions [5]. These dimensions also conveniently align affect, motivation, and behavior in such a way that positive arousal in response to uncertain gains can promote approach behavior, whereas negative arousal in response to uncertain losses instead can promote avoidance behavior (Figure 1) [6,7]. By temporal extension, these 'anticipatory affective' states may promote short-term survival as well as long-term adaptation. Beyond merely neatly summarizing descriptions of affective experience, then, positive arousal and negative arousal dimensions could reflect the underlying activity of distinct brain circuits that generate affective experience and behavior [8].

Challenge and approach

A critical theoretical challenge for affective neuroscience involves identifying neural generators of human affective experience and behavior [9,10]. A method for inferring affect from fMRI data could help researchers to address this challenge. Furthermore, given that practical applications of neuroimaging (e.g., in health or choice) often ultimately seek to infer experiences or behaviors from brain activity, the question may not be so much whether to make inferences, but rather how to make the best inferences. Fortunately, affective space not only implies tools for measuring affective experience, but also a scheme for linking neural activity to affective experience. Based on the geometry of affective space, we propose a quantifiable and testable mapping of fMRI activity onto affective experience and behavior. However, is such a mapping even possible and, if so, how could it be implemented?

Corresponding author: Knutson, B. (knutson@psych.stanford.edu).

Keywords: accumbens; insula; positive; negative; arousal; affect; human; neuroimaging.

1364-6613/

 http://dx.doi.org/10.1016/j.tics.2014.04.006

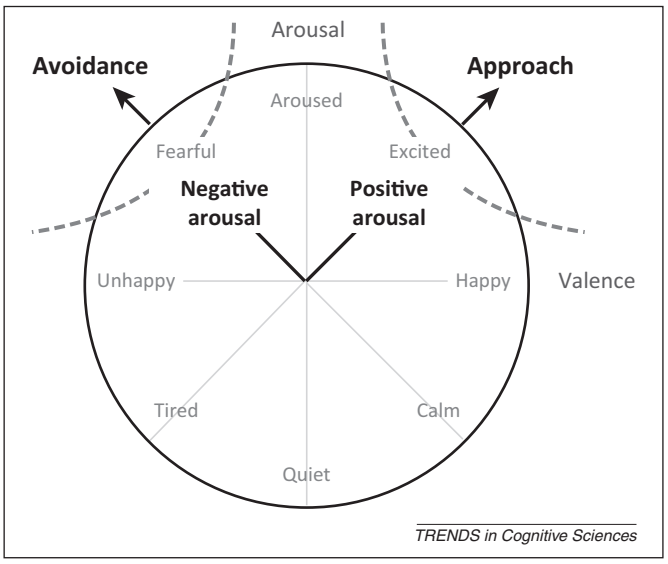

Figure 1. An affective circumplex, emphasizing continua from positive arousal to approach and from negative arousal to avoidance.

Some have argued that mapping brain activity to affective experience constitutes a technically impractical or even conceptually impossible task. Technically, Poldrack [11] suggested that the apparent lack of functional specificity of some brain regions limits inferences that can be drawn from their activity. Affective processes (e.g., arousal) may occur more generally than specific cognitive processes associated with symbolic representation (e.g., language production), and so appear in the context of many different tasks. Furthermore, by reducing the number of brain features and affective features under consideration, investigators can enhance inferential sensitivity and minimize potentially spurious findings. Conceptually, after reviewing failures of past neuroimaging studies to consistently associate human brain activity with emotional responses, Lindquist and colleagues [12] recommended that scientists abandon attempts to link local brain activity to emotional experiences (see also [13]). However, the spatial and temporal resolution of neuroimaging methods and the sophistication of analytic tools continue to improve with each passing year, and previous failures may merely reflect historical limitations in the resolution of designs, acquisition, and analyses.

Others have implicitly or explicitly endorsed affective inference from fMRI data. For instance, Singer *et al.* [14] suggested that insular activity correlates with feelings of general arousal [14], and Paulus and Stein [15] more specifically argued for a role of the insula in the experience of negative arousal. Knutson and Greer [6] further suggested that NAcc activity specifically correlates with positive arousal. Although these claims imply that inferring affect from fMRI data is possible, none have specified exactly how neural markers might combine to support affective inference.

Affective inference from fMRI data would ideally involve data acquisition at the appropriate spatial resolution and a matching temporal resolution. Deep brain stimulation and lesion findings in animals and humans suggest that activity in subcortical circuits generates affective experience and behavior [9,16]. However, many fMRI studies have not implemented sufficiently focused design, acquisition, or analysis protocols to resolve activity in small subcortical circuits [17]. Furthermore, whereas affective states can shift on a second-to-second basis (possibly consistent with changes in neurotransmitter firing rates), most fMRI study designs and analyses lack sufficient temporal resolution to resolve these rapid changes in brain activity. Thus, most existing fMRI studies have not acquired neural data at sufficient spatial resolution or affect data at a matching temporal resolution. These mismatches in measurement may partially account for historically mixed results in studies that seek to infer affect from fMRI data.

To explore links between fMRI activity and affect, we focus here on anticipatory affect, which involves both arousal and valence, and which should ultimately promote approach or avoidance behavior [6]. Many (but not all) studies investigating anticipatory affect utilize monetary incentives, which enable investigators to distinguish between neural responses correlated with affective valence versus arousal. Some of these studies have also assessed brain activity in small subcortical circuits as well as attempted to probe affect on a second-to-second timescale.

Mapping fMRI data to affect

To infer affect from fMRI data, we propose three 'recipes' for: (i) acquiring and preprocessing fMRI data (Box 1); (ii)

Box 1. Extracting fMRI data

Although comprehensive summaries of standard fMRI analyses exist (e.g., [18]), extracting raw data from small subcortical regions requires special considerations that may deviate from typical analysis pathways, as detailed below.

Quality assurance

Given the small subject samples involved and high cost of fMRI, a few outliers can seriously corrupt statistical analyses (e.g., motion or spiking artifacts can introduce outliers that deviate by orders of magnitude). Even after motion correction, we recommend removing data in which the brain moves greater than half a voxel from one volume acquisition to the next, or censoring data that deviate from the average by more than three standard deviations [35].

Derivation of percent signal change

Given that fMRI data produces arbitrary intensity units, only relative rather than absolute changes in activity (against a background of noise) can be inferred. To enhance the comparability of changes in activity between regions, after filtering the data to remove long temporal trends, one can calculate measures of 'percent signal change' by subtracting the average activity across an entire experiment in a given voxel from the activity at a given time point and then dividing by the average activity over time multiplied by 100.

Averaging and extracting activity

Based on previous meta-analyses of brain foci implicated in anticipatory affect (see above), one can specify 8-mm diameter spherical volumes of interest around NAcc (e.g., $x = \pm 10$, $y = 10$, $z = -2$) and anterior insula (e.g., $x = \pm 33$, $y = 23$, $z = -4$) foci in standardized space (e.g., Talairach space [6]). In the event of individual differences in cortical localization, these volumes of interest can be adjusted in the right–left and superior–inferior planes to fit anatomical landmarks based on higher resolution colocalized structural scans (e.g., [36]). Percent signal change can then be averaged within each volume of interest and then peak activity occurring 4–6 s after an event of interest can be selected (e.g., an incentive cue presented during the MID task).

Box 2. Measuring affective responses

The architecture of affective space implies independent dimensions of valence and arousal, which can arbitrarily be rotated. Quality assurance measures might include excluding data from subjects who report no affective variation across an entire range of experimental stimuli. To derive measures of positive arousal and negative arousal, we mean-deviate and rotate valence and arousal ratings as described below:

(i) Mean-deviate valence and arousal ratings across stimuli within each subject;

(ii) Project these ratings onto 45° rotated positive arousal and negative arousal axes as follows: positive arousal = (arousal + valence)/sqrt(2); and negative arousal = (arousal – valence)/sqrt(2).

Arousal and valence probes can be obtained either during an experiment or afterwards, in reference to specific experimental events. Instruction in the meaning and use of the ratings can improve comprehension (i.e., supplement). Acquiring arousal and valence ratings in response to several events for each subject can enhance interpretability and reduce response biases, because the raw ratings are initially converted to relative ratings within subject before transformation [27]. Note that, although these measures implicitly assume that affective experience can be accessed by conscious introspection, they do not assume that affect requires insight (similar to psychophysical studies of visual perception).

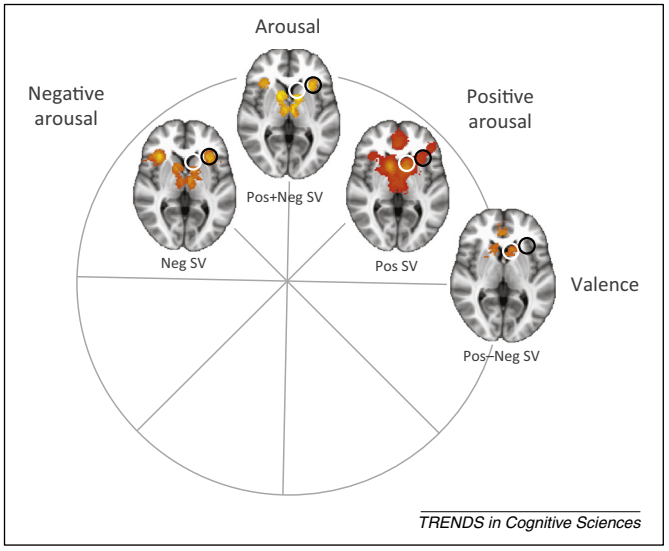

Figure 2. Meta-analytic results for activity in nucleus accumbens (NAcc; white circles) and anterior insula (black circles) during incentive anticipation. Activation likelihood estimate maps adapted from Figure 3 in [21] superimposed onto the affective circumplex [from right to left: positive minus negative subjective value (SV), positive subjective value, positive plus negative subjective value, and negative subjective value].

acquiring and preprocessing affect data (Box 2); and (iii) matching and inferring affect from fMRI data (Box 3). All proposed transformational mapping schemes are informed by the geometry of affective space.

Methods for fMRI data acquisition and analysis continue to evolve. One popular analytic approach involves contrasting fMRI activity from different conditions across the entire brain to identify which regions show differential involvement [18]. Here, we instead adopt a more basic approach of identifying volumes of interest based on previous findings, followed by averaging and extracting raw data. Whereas some meta-analyses of fMRI studies may not have implicated coherent circuits in emotional categories (e.g., [12]), others have clearly implicated a remarkably consistent set of brain regions in incentive anticipation. Specifically, several meta-analyses have implicated the NAcc and sometimes the medial prefrontal cortex (MPFC) in anticipation of gains, but the anterior insula and sometimes medial caudate in anticipation of losses as well as gains [6,19–22] (e.g., Figure 2). Based on these findings, we first extract data from NAcc and anterior insula volumes of interest associated with incentive anticipation to map to anticipatory affect, thereby first reducing the number of features in the neuroimaging data (Box 1; MFPC and medial caudate data were also explored, but either failed to improve or degraded derived solutions, and so were not included).

Neuroimaging researchers understandably often pay closer attention to measurement of brain activity than to measurement of affect. Affect measurement minimally requires assessing valence and arousal (or rotations of these dimensions) in response to specific stimuli or events, either during or after an experiment. As with fMRI activity (Box 1), an individual's relative affective response to different experimental stimuli can be calculated by subtracting their averaged affective response across all items from their affective response to each item. Thus, relative affective impact can be compared in a manner similar to relative regional brain activity in fMRI data (Box 2).

To link brain activity and affect, the most straightforward scheme might involve mapping two neural mechanisms directly onto two independent dimensions of affective space (for instance, NAcc to positive arousal and anterior insula to negative arousal). However, the combined meta-analytic evidence to date suggests a less separable (yet still tractable) mapping [6,19–22]. Whereas NAcc activity includes both valence and arousal components, thus potentially mapping onto positive arousal, anterior insula activity primarily carries an arousal component, thus only partially mapping onto negative arousal. However, given knowledge of the approximate relative position of the two brain activity vectors in affective space (e.g., an approximately 45° offset) and a few simplifying assumptions, these two spaces can be mapped to derive inferences about positive arousal and negative arousal by combining NAcc and anterior insula activity (Box 3; Figure 3).

Testing the mapping

Can the proposed mapping work in theory? Consistent with the anticipatory affect model, when humans playing the monetary incentive delay (MID) task anticipated large uncertain gains, they showed increased NAcc and anterior insula activity and reported experiencing increased positive arousal, but when they anticipated large uncertain losses, they only showed increased anterior insula activity and reported experiencing increased negative arousal [6]. Applying the proposed transformations to changes in fMRI activity typically observed in the MID task (Box 3), if gain cues increase NAcc activity by 0.1% and also increase anterior insula activity by 0.1%, then neurally inferred positive arousal increases by 1 point, whereas inferred negative arousal increases only by 0.20 points (see 'A' in Figure 3).

Box 3. Mapping fMRI data to affect

We seek to infer positive arousal and negative arousal from NAcc and anterior insula activity. Assuming that NAcc activity maps onto positive arousal and that anterior insula activity maps onto general arousal, and multiplying by an arbitrary scaling factor (here, k ~ 10) to scale from percent signal change to rating points suggests the following transformations:

(i) neurally inferred positive arousal = NAcc activity * k;

(ii) neurally inferred negative arousal = (Anterior insula activity – (NAcc activity/sqrt(2))/sqrt(2)) * k

Note that, because anterior insula activity maps onto general rather than negative arousal, a component of NAcc activity is subtracted from anterior insula activity to derive neurally inferred negative arousal. Thus, the transformations remap neural activity on valence and arousal dimensions to positive arousal and negative arousal dimensions via projection. Although a 45° rotation is initially assumed, angular parameters could be empirically tested and modified on the basis of future findings to accommodate greater or lesser degrees of rotation (e.g., [37]). If brain and self-report data are acquired at a similar temporal resolution, the fit of neurally inferred affect to self-reported affect could be compared across experimental conditions, either relative to chance or an existing benchmark (e.g., using classification and cross-validation techniques).

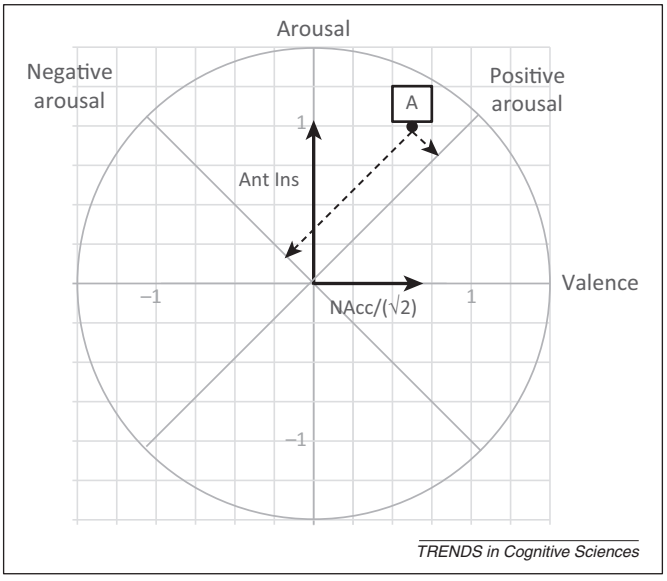

Figure 3. Neurally inferred positive arousal and negative arousal from nucleus accumbens (NAcc) and anterior insula (Ant Ins) activity.

Do the proposed transformations map onto experimental findings? Data drawn from 12 healthy young subjects who completed the MID Task (age range = 20–50 years) provided both cue-elicited brain activity and cue-elicited affect ratings (data for this subsample were drawn from a larger published lifespan sample described in [23]). During the MID task, subjects saw cues indicating that they could gain or lose US$0.00, US$1.00, or US$5.00 by responding to a subsequent rapidly presented target with a button press. fMRI activity was extracted from NAcc and anterior insula volumes of interest 6 s after the presentation of each cue and averaged by cue type, whereas affect ratings in response to each cue type were acquired immediately following the experiment.

Replicating patterns of brain activity previously observed in the MID task [6,21], anticipation of large gains increased NAcc activity (relative to the nonincentive conditions; $P < 0.01$), whereas anticipation of both large gains and large losses increased anterior insula activity relative to the nonincentive conditions ($P < 0.05$). For affective ratings, anticipation of large and medium gains increased positive arousal relative to the nonincentive conditions ($P < 0.01$), whereas anticipation of large and medium losses increased negative arousal relative to the nonincentive conditions ($P < 0.05$). fMRI activity was then transformed into neurally inferred affect (using the functions described in Box 3). For neurally inferred affect, anticipation of large gains increased inferred positive arousal relative to the nonincentive conditions ($P < 0.01$), whereas anticipation of large losses increased inferred negative arousal relative to the nongain ($P < 0.05$) but not the nonloss condition, approximating patterns observed in affect ratings (i.e., black versus gray lines in Figure 4). Finally, across conditions, nonparametric permutation tests (n = 100 000 repetitions) revealed a significant association of neurally inferred positive arousal with rated positive arousal (slope = 0.58, $P < 0.05$), as well as a trend towards an association of neurally inferred negative arousal with rated negative arousal (slope = 0.29, $P < 0.10$) across conditions (Figure 4).

Concluding remarks and implications

The proposed mapping of fMRI data to affect produces interpretable initial results when applied to ideal as well as actual group data. The goal of this preliminary mapping is not to provide a comprehensive or final scheme, but rather an initial step towards inferring affect from fMRI data. Here, rather than pursuing a 'broad' account across any single level of analysis (e.g., chemistry, fMRI activity, affect, or motivated behavior), we instead seek to establish a few 'deep' links between adjacent levels of analysis (i.e., fMRI activity and affect). Once established, these links can then be enriched and elaborated in terms of breadth within each level of analysis as well as in terms of depth across multiple levels of analysis (e.g., Figure 5). If the proposed links replicate and generalize, they could lay a foundation for inferring affect from fMRI data. However, validating such a mapping first requires a framework that generates quantifiable, directional, and testable predictions.

The relatively straightforward mapping from NAcc activity to positive arousal has been described in several studies (reviewed in [6]). This mapping suggests that affect arises during anticipation of uncertain incentives, and not just in response to incentive outcomes, therefore moving beyond consequentialist theories by indicating that neural mechanisms generate affect during anticipation as well as in response to outcomes [24]. Given that anticipatory affect can occur before choice, associated neural markers may best predict eventual choice. The link between subcortical activity in the NAcc and positive arousal further implies that some affective states may be conserved across mammalian species and development, and need not invoke reflective processes associated with higher cortical input (e.g., [13]). Consistent with this line of reasoning, dopamine (but not norepinephrine) selectively innervates anterior parts of the NAcc and its synaptic availability fluctuates on a second-to-second basis. Dopamine release has been speculated to increase NAcc fMRI activity [25], a prediction that can now be directly evaluated with new neuroscience tools (e.g., optogenetics [26]).

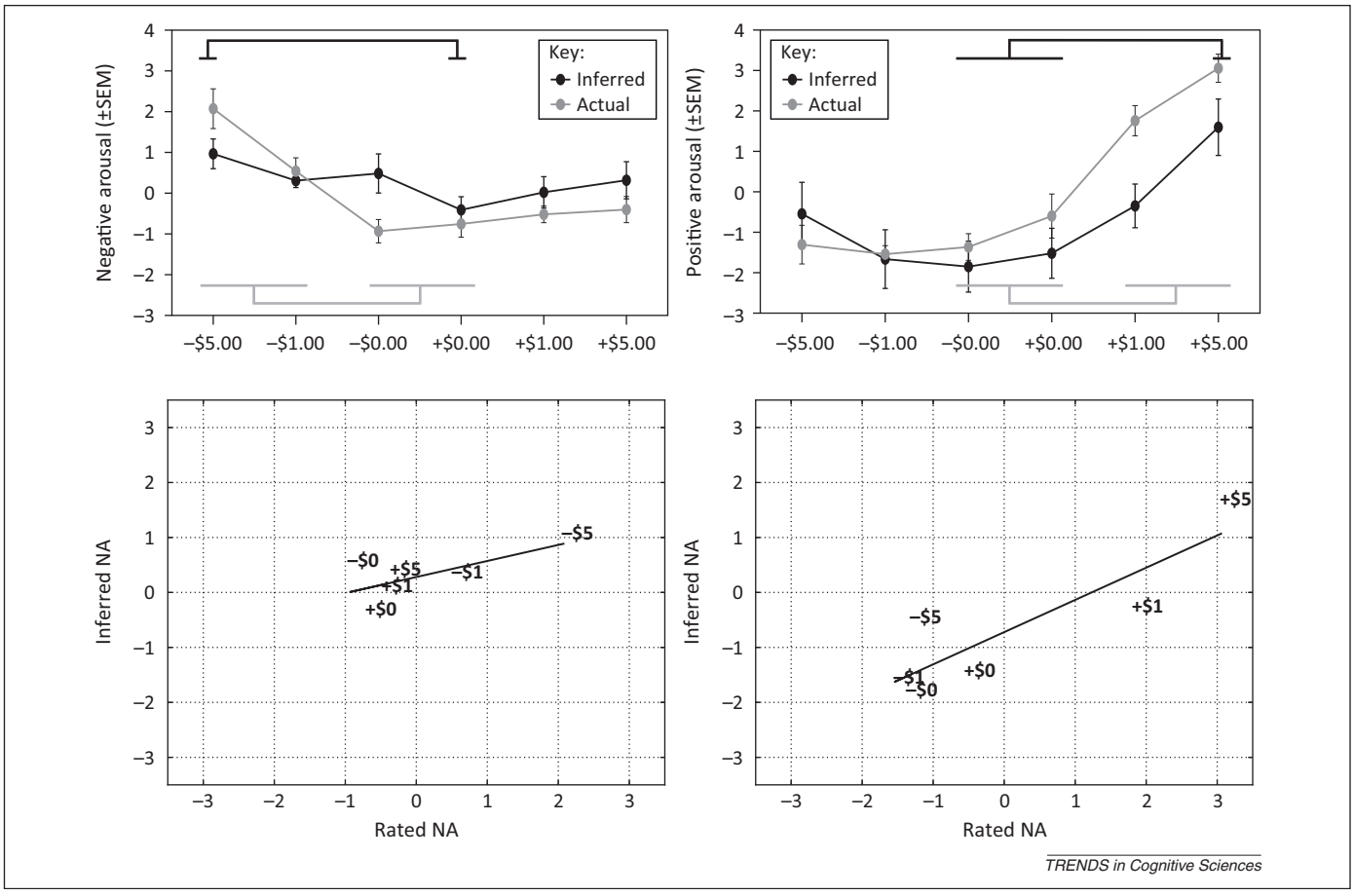

Figure 4. Neurally inferred versus rated affect ($n = 12$; mean ± SEM; lines indicate significantly different conditions), and associations of neurally inferred with rated affect across incentive conditions. Abbreviations: NA, negative arousal; PA, positive arousal.

The mapping from anterior insula activity to negative arousal is less straightforward. As suggested by meta-analytic findings (and to the surprise of many researchers), localizing fMRI activity strictly associated with negative arousal has proven more elusive than that associated with positive arousal. This more ambiguous mapping may result from either conceptual factors (e.g., greater diversity in highly arousing negative emotions) or physiological factors (e.g., involvement of more diverse neurotransmitters and target regions), or both. Regardless, by assuming partial independence of the projections of anterior insula and NAcc activity onto affective space, fMRI activity from these regions might still support inferences about negative arousal (Figure 5). Unlike the anterior NAcc, the anterior insula is innervated by both dopamine and norepinephrine, both of which likely vary from second to second. Extending the previous logic, release of both dopamine and norepinephrine may account for increased but nonspecific anterior insula fMRI activity observed during anticipation of gains and losses [6].

Activity from other brain regions commonly associated with affective responses (e.g., amygdala or prefrontal cortex) do not as reliably appear in meta-analyses of fMRI activity during incentive anticipation, and activity extracted from these regions did not improve affective inferences using the current scheme, but nonetheless deserve future exploration. Other small subcortical regions may have more direct roles in generating negative arousal, but have historically been difficult to visualize with fMRI (e.g., the ventromedial hypothalamus and periaqueductal gray), and these also deserve further scrutiny.

The proposed mapping raises a host of possibilities for development. With respect to fMRI activity, activity from other regions (mentioned above) could be added to augment existing predictions. For instance, regions involved in value integration (e.g., the medial prefrontal cortex [27]) or reflection (e.g., the dorsal medial prefrontal cortex [28]) might directly or indirectly influence links between fMRI activity and affect. New multivariate techniques could also be used to determine whether patterned activity or altered connectivity provides additional information about affect (e.g., [29]). Spatial and temporal noise can corrupt fMRI data, and so analyses using new feature selection algorithms with noise penalties might improve neural inferences about affect (e.g., [30]). Beyond group inference, the power of the proposed scheme to illuminate individual differences could also be explored. Perhaps most importantly, the proposed mapping must be generalized to other scenarios (e.g., affective responses to outcomes), and tasks that evoke less specific and controlled shifts in affect.

Opportunities also exist for improving the measurement of affect. Assessment of self-reported affect often requires trading semantic coverage (e.g., number of items) against temporal precision (e.g., density of items). By using dimensional probes, investigators can briefly survey most of affective space with just two strategically chosen items

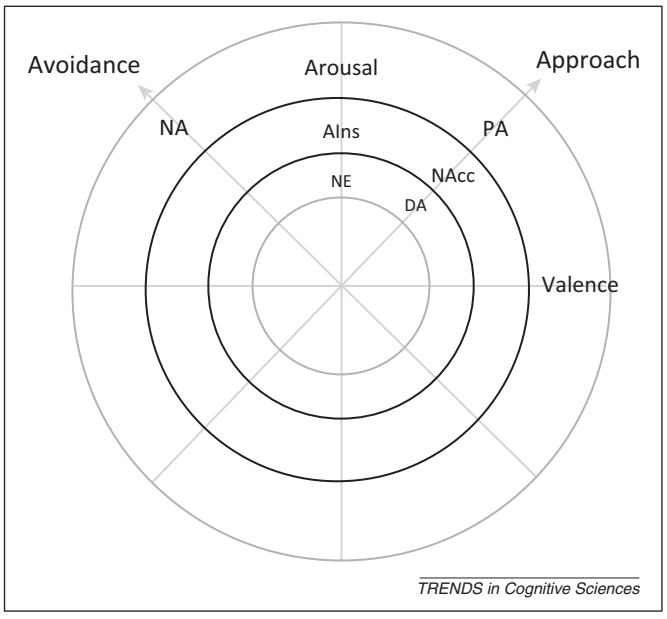

Figure 5. Proposed mapping of constructs across levels of analysis (ascending levels from inner to outer include neurochemistry, functional magnetic resonance imaging activity, affective experience, and motivated behavior). Abbreviations: AIns, anterior insula; DA, dopamine; NA, negative arousal; NAcc, nucleus accumbens; NE, norepinephrine; PA, positive arousal.

(see also [31]). Interrupting subjects during an affectively engaging task, however briefly, may still alter existing affect. Furthermore, humans may fail to accurately report past affective experiences for several reasons, including but not limited to: failing to pay attention; forgetting what was experienced; reporting a different intervening affective state; reconstructing experience based on scripts or rules; being unwilling to report socially undesirable experiences; or a simple lack of reflective capacity. Fortunately, rapidly probing affect in response to relevant events can minimize some of these problems, as can comparing affective responses across several different events. Perhaps most importantly, investigators should seek to match the temporal specificity of affect probes to that afforded by fMRI data.

Although the current scheme links brain activity in a constrained set of regions to affect, it does not yet extend to dynamic changes in affect or more specific emotions. However, once static mappings are established, they could be temporally extended to track 'affect dynamics' online [32], which might provide information about the trajectory and duration of affective episodes. Whereas researchers have traditionally thought of emotions as points in affect space, affect dynamics might provide richer tools for describing affective or emotional episodes (e.g., sufficient movement up and to the left might imply anxiety, movement straight to the left might imply anger, and movement down and to the left might imply sadness). Affect dynamics could also inform researchers about not only momentary changes in affect, but also individual differences in affective trajectories over time, with potential implications for diagnosing and tracking psychiatric symptoms.

Mapping fMRI activity to affect could yield both theoretical and practical benefits. Theoretically, such a mapping could make affective inference from fMRI data more testable and quantitative. Mappings may also resolve currently conflicting interpretations of fMRI data. For instance, researchers have debated about whether NAcc activity reflects valence or salience (which could be reframed as arousal; e.g., [33]). The current scheme implies that NAcc activity reflects both valence and arousal, and additionally suggests the novel prediction that NAcc activity can vary independently of negative arousal (e.g., feelings such as anxiety and tension). Furthermore, because fMRI activity changes from second to second, affective inference from fMRI activity might offer a glimpse into dynamics of affect that have historically eluded measurement. One interesting challenge involves the possibility that fMRI activity might in some cases provide a more accurate index of affective responses than self-report. Given that positive arousal aligns with approach behavior and negative arousal aligns with avoidance behavior, affective inference should extend not only to self-reported affective experience, but also to eventual behavior. Practically, mapping fMRI activity to affect could complement existing self-report measures of affective experience. As people primarily visit psychiatrists due to a lack of excitement or an excess of distress, these tools might prove useful for predicting psychiatric diagnoses and monitoring therapeutic progress. Additionally, because subjective reactions rather than objective perceptions drive behavior, these tools may prove useful for understanding which features of a proposition most powerfully motivate choice (e.g., to make a purchase or investment). Thus, principled inferences could help realize the potential of affective neuroscience to improve human health and well-being.

For Galileo, matching the resolution of his method (the telescope) to the phenomenon of interest (the movement of Jupiter's moons) yielded a scientific breakthrough. Although scientists do not yet know the best resolution for linking brain activity and affect, research suggests that second-to-second changes in the activity of small subcortical circuits powerfully and unconditionally elicit affective behavior [34]. By matching the resolution of neural and affective measures, researchers can begin to map links between brain activity and affect in humans. In the future, scientists may track affect dynamics with the same facility that astronomers traced the trajectories of moons through outer space, with similarly revolutionary consequences for our understanding of inner space.

Acknowledgments

We thank Christian Buchel, Jeanne L. Tsai, Charlene C. Wu, and two anonymous reviewers for suggestions on previous drafts, as well as Jonathan E. Taylor for statistical consultation. This work was supported by a FINRA Investor Education Foundation Grant and a Bio-X seed grant to B.K.

References

1 Wundt, W. (1897) *Outlines of Psychology*, Wilhelm Engelmann
2 Russell, J.A. (1980) A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178
3 Yik, M.S.M. *et al.* (1999) Structure of self-reported current affect: integration and beyond. *J. Pers. Soc. Psychol.* 77, 600–619
4 Watson, D. and Tellegen, A. (1985) Toward a consensual structure of mood. *Psychol. Bull.* 98, 219–235
5 Bradley, M.M. *et al.* (2001) Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion* 1, 276–298

6 Knutson, B. and Greer, S.M. (2008) Anticipatory affect: neural correlates and consequences for choice. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 363, 3771–3786
7 Larsen, J.T. *et al.* (2001) Can people feel happy and sad at the same time? *J. Pers. Soc. Psychol.* 81, 684–696
8 Watson, D. *et al.* (1999) The two general activation systems of affect: structural findings, evolutionary considerations, and psychobiological evidence. *J. Pers. Soc. Psychol.* 76, 820–838
9 Panksepp, J. (1982) Toward a general psychobiological theory of emotions. *Behav. Brain Sci.* 5, 407
10 Davidson, R.J. and Sutton, S.K. (1995) Affective neuroscience: the emergence of a discipline. *Curr. Opin. Neurobiol.* 5, 217–224
11 Poldrack, R.A. (2006) Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* 10, 59–63
12 Lindquist, K.A. *et al.* (2012) The brain basis of emotion: a meta-analytic review. *Behav. Brain Sci.* 35, 121–143
13 LeDoux, J. (2012) Rethinking the emotional brain. *Neuron* 73, 1052
14 Singer, T. *et al.* (2009) A common role of insula in feelings, empathy and uncertainty. *Trends Cogn. Sci.* 13, 334–340
15 Paulus, M.P. and Stein, M.B. (2006) An insular view of anxiety. *Biol. Psychiatry* 60, 383–387
16 Coenen, V. *et al.* (2011) Cross-species affective functions of the medial forebrain bundle-implications for the treatment of affective pain and depression in humans. *Neurosci. Biobehav. Rev.* 35, 1971–1981
17 Sacchet, M.D. and Knutson, B. (2012) Spatial smoothing systematically biases the localization of reward-related brain activity. *Neuroimage* 66C, 270–277
18 Huettel, S.A. *et al.* (2004) *Functional Magnetic Resonance Imaging*, Sinauer Associates
19 Liu, X. *et al.* (2011) Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 35, 1219–1236
20 Diekhof, E.K. *et al.* (2008) Functional neuroimaging of reward processing and decision-making: a review of aberrant motivational and affective processing in addiction and mood disorders. *Brain Res. Rev.* 59, 164–184
21 Bartra, O. *et al.* (2013) The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76, 412–427
22 Clithero, J.A. and Rangel, A. (2013) Informatic parcellation of the network involved in the computation of subjective value. *Soc. Cogn. Affect. Neurosci.* http://dx.doi.org/10.1093/scan/nst106
23 Wu, C.C. *et al.* (2013) Affective traits link to reliable neural markers of incentive anticipation. *Neuroimage* 84C, 279–289
24 Loewenstein, G.F. *et al.* (2001) Risk as feelings. *Psychol. Bull.* 127, 267–286
25 Knutson, B. and Gibbs, S.E.B. (2007) Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology* 191, 813–822
26 Witten, I.B. *et al.* (2011) Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* 72, 721–733
27 Knutson, B. *et al.* (2005) Distributed neural representation of expected value. *J. Neurosci.* 25, 4806–4812
28 Fleming, S.M. and Dolan, R.J. (2012) The neural basis of metacognitive ability. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 367, 1338–1349
29 Wager, T.D. *et al.* (2013) An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* 368, 1388–1397
30 Grosenick, L. *et al.* (2013) Interpretable whole-brain prediction analysis with GraphNet. *Neuroimage* 72, 304–321
31 Russell, J.A. *et al.* (1989) Affect grid: a single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* 57, 493–502
32 Nielsen, L. *et al.* (2008) Affect dynamics, affective forecasting, and aging. *Emotion* 8, 318–330
33 Brooks, A.M. and Berns, G.S. (2013) Aversive stimuli and loss in the mesocorticolimbic dopamine system. *Trends Cogn. Sci.* 17, 281–286
34 Panksepp, J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions*, Oxford University Press
35 Wilcox, R.R. (1998) How many discoveries have been lost by ignoring modern statistical methods? *Am. Psychol.* 53, 300–314
36 Samanez-Larkin, G.R. *et al.* (2007) Anticipation of monetary gain but not loss in healthy older adults. *Nat. Neurosci.* 10, 787–791
37 Suri, G. *et al.* (2013) Predicting affective choice. *J. Exp. Psychol. Gen.* 142, 627–632