



(2000). *Neuropsychanalysis*, 2:158-162

[The Mechanism of Confabulation Commentary by Marcel Kinsbourne](#)

[Marcel Kinsbourne, M.D.](#)

Three Questions about Confabulation

Which brain structures are impaired in patients who confabulate? By what mechanism is the confabulation generated? What determines the content of the confabulation? DeLuca concentrates on the first question, and Solms on the last, with little overlap. I hope to find a common thread between them by addressing the middle question.

DeLuca offers a comprehensive discussion of the localization of brain pathology in confabulation. He concludes that confabulation is apt to arise when there is conjoint frontal (specifically ventromedial) dysfunction and impaired memory (due to basal forebrain, diencephalic, or mesial temporal damage). This conclusion is persuasive for the bulk of confabulation, which refers to events in the past, and on which DeLuca focuses. Confabulating about the present, and even the future, which is also well documented in the more severe cases, may not require defective memory. Does orbitofrontal dysfunction suffice to explain it?

Following structural lesions, disordered frontal function and impaired memory are constant and persistent, but confabulation comes and goes. Confabulators only sometimes confabulate, and only on some topics. Confabulations may be offered spontaneously, but more often they are uttered only when they are provoked by leading questions. The suggested mechanisms for confabulation that DeLuca briefly summarizes, impaired self-monitoring and strategic retrieval (Moscovitch and Mello, 1997), may be too much of an explanation for the syndrome. They suggest that when memories are imprecise, it is normal to generate confabulatory material up to the point of utterance, and that it is self-monitoring that warns the individual not to speak but to try again, or admit ignorance. If the reasons for confabulation are so general, then why do even severely affected patients not confabulate during much or all deliberate activity, but only periodically? There must be additional factors that interact with the neuropsychological risk factors to generate overt confabulatory responses.

Factors That Predispose to Confabulation

Why confabulate rather than say that one does not know the answer? I suggest that one contributory factor is the affective significance of the topic about which the patient confabulates. Patients mostly confabulate about personal matters that are emotionally important to them, such as the integrity of their body and their prospects for recovering and for reassuming the prior lifestyle and employment. This is particularly obvious in the patient with a major handicap of which he is totally unaware.

In unilateral neglect, which does not involve frontal injury, the confabulations are not spontaneous, but are narrowly targeted responses to clinicians who tell the patients that they have a disability. They are clearly, albeit unconsciously, defensive. To explain this, it does not suffice to implicate unacceptable grief at the loss or "death" of a limb. Such crippling of body parts is not infrequent, and in the absence of posterior parietal disease with neglect, elicits no such reaction. The key is the bias in attention, which leaves the patient in an irresolvable conflict. The patient who

Dr. Kinsbourne is Professor of Psychology, New School University, New York City.

- 158 -

cannot attend to the left side of things, including his own body, is not only unaware of the disability, but feels physically intact, because there is nothing amiss where his biased attention is focused. By definition, one cannot attend to one's own inability to attend. The feeling of intactness is threatened by conflicting evidence for the left-sided impairments. This conflict between usually reliable sources of evidence subjects the patient to an experiential crisis, an internal contradiction between what the patient feels and what he knows (Kinsbourne, 1987). The patient denies that his left arm is paralyzed and derogates the patently useless limb, which he can see but which he does not feel he owns, and rejects it as belonging to another person. The confabulation is a psychodynamic reaction to an organic problem. As Solms (this issue) remarks, there is an element of wish fulfillment in the content of the confabulation. However, this does not make it qualitatively different from mistaken remembering by people with intact brain function, whose memories are well known to incorporate somewhat self-serving extrapolations. When confabulation is extreme, it appears quite bizarre, and patently at odds with the patient's life situation. What is interesting is that the patient is so free and even prolific with his responses, and so at ease with the contradictions, or willing to modify his confabulation so readily to accommodate contradictions in a superficial manner. When the patient with impaired memory is asked to remember an event, and confabulates about it, what he confabulates, though it is at variance with the facts, is not off topic. So something is being remembered, rather than nothing (Brown, 1988). Emotionally it is on target. Typically the confabulation implies integrity of the body, preserved cognitive function, and freedom of action. It overtly or implicitly denies disabilities of which the patient is not fully aware, and which he has not been able to integrate into his self-image. The confabulations may extend into relationships within the patient's family, so as to cast a more favorable light on difficult family interactions (Conway and Tacchi, 1996). What is it about the combination of poor memory and ventromedial frontal impairment that evokes voiced or acted out wish fulfillment? Why do people with this combination of deficits make statements and perform actions that are obviously incoherent, when they are in no other respects confused?

A failure of self-monitoring is usually suggested. But this does not explain why the confabulated material is generated in the first place, so as to need monitoring. It is not plausible that people generally entertain fantasies when asked straightforward questions, and only fail to voice them because an indwelling monitor censors them. That there is a monitor in the brain that supervises other parts of the same brain is

itself a dubious homuncular concept. Like many other influential theorists, Freud embraced this concept explicitly, conceiving of the ego as an inner eye that, to the extent possible, surveys the id. We can now discard this Cartesian notion (see Kinsbourne, 1988; Dennett and Kinsbourne, 1992). Instead, I suggest that confabulation is fostered by a particular state of mind. This is an inner-directed focus on an affectively laden issue, a focus that is so intense and narrow that it excludes peripheral information, or memories that might conflict with the favored interpretation of the situation. I conceive of this state of mind as being at one extreme of a dimension of states of mind that ranges between extreme interactivity and extreme intra-activity (Kinsbourne, 2000a, in press, a).

The Interactive-Intra-Active Continuum

Much of what people do engages the body and the external environment with action-perception and anticipation-preparation loops through brain, body, and ambient space (Kinsbourne, 2000b). This happens particularly when one is engaged in an effortful interaction, such as a sport or other strenuous occupation. In such interactive circumstances, one is fully engaged in the reciprocal exchange with the person or object in the world, and has no attention left over to introspect. On other occasions, the environment, ambient and bodily, makes no demands at all, and one is free to “go internal.” In fact, one has little choice, as the traffic of the living brain is never ending, and its subjective aspect is thought, and so thought must go on even when there is no occasion for action. Varieties of deep thinking and imaging, such as dreams and reveries, are fully intra-active. These mental states are entirely internally generated. Most of the time, one's mind shuttles between these polar states, partly interacting but introspecting at the same time, as in a thoughtful conversation or when reading a thoughtprovoking book. People move readily between these states, unless they are newborns, who have no choice but to be fully interactive, or people with autism, who are stuck at the intra-active extreme. How are these opposing states instantiated in the brain?

Pribram (1975) remarked “on a currently neglected aspect of brain function—its spontaneous activity, its generative capacities.” Pribram likened this “change in view” to the change in chemistry from

- 159 -

analyzing simple one-way reactions to analyzing reciprocally interacting thermodynamic systems. A quarter of a century later, this advanced paradigm has still not caught on among brain scientists. Yet the organization of the cortical network clearly manifests the reciprocal relationships between the “central processor,” and the input and output mechanisms, that Pribram described. The cerebral gray matter is a recursive network that is largely composed of sequentially arranged processing units, some related to sensation and some to motor control. They constitute the bulk of the neural tissue between the brain's core and its interface with receptor surfaces and effector mechanism (Pandya, Seltzer, and Barbas, 1988). Cerebral areas, such as the components of the visual dorsal and ventral streams, interconnect reciprocally, and waves of activation travel in both directions. What is the implication of the almost equally plentiful two-way traffic? I have proposed that the cerebrum is the site of the clash of opposing waves of neuronal activation, centripetal versus centrifugal. Signals transmitted to the

cerebrum from the receptor surfaces initiate centripetal flow of neuronal activation. This is the traditionally described route by which external change impresses itself upon the spontaneous activity of the brain, which responds to that perturbation either by habituating to it or by formulating appropriate action plans to exploit or escape from the changed situation. This direction of flow earned the series of visual areas from V1 to inferotemporal, and from V1 to posterior parietal, the designations ventral and dorsal stream respectively. However, a stream flows in one direction, yet the series of cerebral areas are not unidirectional in their interactions. These streams flow in both directions. The conversely directed centrifugal flow, through the same processing units, but connecting different layers of cortex (5 and 6 with 1, as distinct from 3 and 4 with 4 for the corticopetal), is anchored in limbic cortex. It confronts the input with endogenously developed expectations and evaluations, perhaps by tuning the sensory channels in line with the attributes of the expected stimulus.

The Contents of Intra-Active Mental States

Only inputs that come as a total surprise perturb the brain, which is caught unprepared. More usually, the individual formulates anticipations about what will happen next, and prepares corresponding responses, which turn out to be roughly appropriate. These anticipations are images that approximate the expected event. Anticipations are most fully specified when the individual is interacting in a limited arena. Sports such as tennis, Ping-Pong, or baseball, in which the degrees of freedom are limited, are examples. In more commonly occurring situations, the anticipations are necessarily less specific and differentiated; they are generic. For example, a meeting room or classroom is expected to contain tables, chairs, people, but not lizards and giant artichokes, but without definition as to which of the occupants is where, and how they look. The anticipations have enough degrees of freedom to accommodate such unpredictable variation in detail. In the opposite extreme, anticipations are freewheeling when the individual is not interacting with the exterior at all, but is introspecting, in a reverie, fantasizing or dreaming. Unconstrained by the environment, the selection of thought and imagery will quite normally reflect the person's current motivations, drives, desires, and wishes. It excludes any implications of what is undesired and consequently denied. In short, the centrifugal system attains a variable end point of differentiation, congruent with external reality, or falling short of that, depending on whether the mental state is interactive or intra-active. When the concrete circumstances fail to constrain the centrifugal system, it becomes a vehicle of fantasy and desire.

Frontal Control of the Content of an Utterance

To remember a past episode, one has to detach attention from the here-and-now (Kinsbourne and Wood, 1975). This frees the centrifugal system to offer up memories as reconstructions of past events. To the extent that the memories are ill defined, either because they have long faded, or because the amnesic individual has impaired episodic memory, there is scope to shape or distort the memory in line with the motives of the moment. This process offers the psychoanalytic theorist opportunities for interpretation, as Solms vividly illustrates. Ventromedial frontal lesions reduce the extent to which external reality constrains the contents of fantasy, and a similar opportunity for distortion arises. This is why confabulators

only confabulate at times; namely, those times when their emotions are aroused. At such times the confabulations are wordy, fluid, and inconstant, because they are unconstrained, and do not commit the individual to specific actions (rather like dreams), and they are readily manipulated by suggestion.

- 160 -

Why does ventromedial frontal impairment favor intra-active thinking and a loosening of ambient constraints? By virtue of its dual control of underlying limbic system and of posterior cortical exteroception, the ventromedial frontal lobe invests the evolving thought or intention with relevant context (Nauta, 1964), both from the immediate environment (“there is a policeman at the corner”) and the knowledge base (“getting caught means jail”). When the VNF control over behavior is weakened by brain disease, the approach toward the goal of the drive state remains unqualified by cost-benefit accounting, and an impulsive act results, or a confabulation that is unqualified by logic or contextual reality. I see the constraining role of ventromedial frontal activity as equivalent to what Freud postulated as “binding.” Its absence is, as Luria remarked (as cited by Solms), “a disturbance of the selectivity of mental processes.”

The effects of ventromedial frontal damage stand in opposition to those of damage to left dorsolateral prefrontal cortex. The utterances of patients whose left dorsolateral frontal cortex is damaged are sparse, halting, and preservative; that is, overconstrained. Orbitofrontal damage releases and dorsolateral frontal damage restrains spontaneity. The intact orbitofrontal-dorsolateral opponent system provides a flexible range of expression, between concrete and specific, and figurative and profuse. In imbalance, maladaptive extremes become manifest.

The Unconscious

Confabulations are more easily intertwined in a memory than in a contemporaneous percept, since the former is less clear-cut and constrained by the here-and-now. However, when the emotional drive is intense, even percepts are distorted, and the functional status of memory mechanisms becomes irrelevant. Solms offers examples: the patient who sees the No Smoking sign as a clock face that indicates 5:00 P.M. (visiting hour); the woman who confabulates that the man in the next bed is her husband. In such cases the prefrontal lesion relaxes the extent to which the external reality constrains the experience as it evolves into awareness.

The confabulator is the magical realist of neuropsychology. Most of the time he behaves like everyone else. But when the issue comes close to his emotional core, he takes flight into a magical, self-serving, effortlessly dreamlike solution.

Its customary disconnection from action manifests the inner-directed, self-serving nature of confabulation. The patient who claims good health does not head for the desk to ask to be discharged. The patient who rejects his left arm as being some else's does not spontaneously complain about it and does not push it away. Like a dream, the confabulation is a script that fulfills a need, or dramatizes a fear. It is rarely a prescription for action. It is not a deliberate deception, nor an attempt to conceal a memory problem (Talland, 1965).

This formulation is consistent with Solms's content analysis of confabulation, the special characteristics of which he likens to those attributed by psychoanalytic theory to the system unconscious. It is, of course, not invariably the pleasure

principle that determines the content of intra-active thought processes. The extreme reverse occurs in the “hypofrontal” schizophrenic, who is beset by “fear-fulment” in his toxic hallucinations. Nightmares are a normative example. It is consistent with Solms's reflection that “mature cognitive functions are built on the foundations of more primitive mental functions (that) persist beneath the behavioral surface and continue to exert an effect on adult mental life.” It differs from the psychoanalytic view in that it does not posit an unconscious thought or image that escapes repression so as to enter the conscious sphere. Images are as conscious as are percepts, primary process is as conscious as secondary process. They are alternate states of awareness, externally and internally driven respectively. There is no reason to suppose that primary process imagery preexists in essentially the same form in an unconscious arena before it leaks into the light of awareness. Consciousness is not a place into which images enter, but an attribute of a stage of their formation, namely that stage during which they gain final definition (Brown, 1988). When a representation becomes integrated into the cortical neuronal field, it changes character by so doing (Kinsbourne, 1988). What characterizes “unconscious awareness” is a particular pattern of vaguely sensed motivation, which shapes the conscious experience into primary process form. I consider the unconscious to be a set of motivated predispositions, not a store of full-fledged images and thoughts. These predispositions are most apt to control experience during intra-active states. In confabulation, there is an imbalance between the two directions of flow of neuronal activity, in favor of the centrifugal.

The attributes of the unconscious are those of unconstrained intra-activity. The tolerance of contradiction

- 161 -

derives from the narrow focus on the desired outcome. The narrow focus excludes all else, supportive or not, and relies on the force of conviction alone. Blind faith, it cannot be shaken by reasoned argumentation. The prominent temporal dislocations in confabulated recollections seem to me not to be due to faulty shuffling of time slices by the impaired brain, but to the predominance of the desired memory over the memory that was requested or called for by the situation. Emotional authenticity takes precedence over temporal accuracy and logical coherence. When the clinician points out inconsistencies, the patient brushes them aside. The desired memory may even be for an event that could have occurred, but never has. Kinsbourne and Wood (1975), when they introduced the notion that the amnesic syndrome is a deficit in episodic memory, showed that the “recollections” of amnesics were not so much temporally displaced, but unconsciously manufactured from their knowledge base of what is apt to happen under the specified circumstances.

Subjectivity and Neuroscience

Neuroscience is ambivalent about subjective report. Reluctantly, it relies on subjective report when it comes in quantitative form, as in psychophysics. But when they deal with brain disorders, neuropsychologists are more comfortable with performance deficits than with deviant experience. The productive consequences of brain pathology are paid little attention in texts. The emphasis is always on deficient performance. This is a damaging prejudice. Whereas another person's subjective

state cannot be directly observed, it can to a large degree be communicated, and one can test it both for internal coherence, and for replicability across subjects. Cognitive neuroscience handicaps itself when it treats subjective reports by brain-damaged people cavalierly and focuses on only part of the story, the negative symptoms. The positivistic undervaluing of subjectivity by neuropsychologists is a vestige of Cartesian dualism, which exalts conscious content (the “Cogito”), but relegates it to scientific limbo. The subjective aspect of the mind is no less a product of the brain than observable behavior; mental states are brain states. The subjective report is a source of insight into how the brain works (Brown, 2000).

References

- Brown, J. W. (1988), *The Life of the Mind*. Hillsdale, NJ: Erlbaum.
- Brown, J. W. (2000), *Mind and Nature: Essays on Time and Subjectivity*. London: Whurr.
- Conway, M. A., & Tacchi, P. C. (1996), Motivated confabulation. *Neurocase*, 2:325-339.
- Dennett, D., & Kinsbourne, M. (1992), Time and the observer: The where and when of consciousness in the brain. *Behav. & Brain Sci.*, 15:183-247.
- Kinsbourne, M. (1987), Mechanisms in unilateral neglect. In: *Neurophysiological and Neuropsychological Aspects of Spatial Neglect*, ed. M. Jeannerod. Amsterdam: Elsevier.
- Kinsbourne, M. (1988), Integrated cortical field model of consciousness. In: *The Concept of Consciousness in Contemporary Science*, ed. A. J. Marcel & E. Bisiach. Oxford: Oxford University Press.
- Kinsbourne, M. (2000a), New models for old: Taking the neural network seriously. *Brain & Cognit.*, 42:13-16.
- Kinsbourne, M. (2000b), Consciousness in action: Antecedents and origins. *Mind & Lang.*, 15:545-555.
- Kinsbourne, M. (in press, a), Dynamic self-organization of the cerebral network. In: *On the Dynamics of Modern, Complex, Social, and Democratic Systems*, ed. J. Gotschl. Dordrecht, Netherlands: Kluwer.
- Kinsbourne, M., Wood, F. (1975), Short term memory processes and the amnesic syndrome. In: *Short Term Memory*, ed. J. A. Deutsch. New York: Academic Press.
- Moscovitch, M., & Melo, B. (1997), Strategic retrieval and the frontal lobes: Evidence from confabulation and amnesia. *Neuropsychologia*, 35:1017-1034.
- Nauta, W. (1964), Some efferent connections of the prefrontal cortex in the monkey. In: *The Frontal Granular Cortex and Behavior*, ed. J. Warren & K. Ackert. New York: McGraw-Hill, pp.
- Pandya, D., Seltzer, B., & Barbas, H. (1988), Input-output organization of the primate cerebral cortex. In: *Comparative Primate Biology*, ed. H. D. Stecklis & J. Erwin. New York: Liss, pp.
- Pribram, K. H. (1975), The primate frontal cortex: Progress report 1975. *Acta Neurobiol. Exp.*, 35:609-625.
- Talland, G. A. (1965), *Deranged Memory*. New York: Academic Press.

Kinsbourne, M. (2000). The Mechanism of Confabulation Commentary by Marcel Kinsbourne. *Neuro-Psychoanalysis*, 2:158-162