The psychological unconscious and the self

John F. Kihlstrom

The Amnesia & Cognition Unit, Department of Psychology, University of Arizona, Tucson, AZ 85721. USA

Abstract. Documentation of implicit expressions of memory in head-injured, mentally ill and normal individuals has offered a new perspective on the problem of unconscious influence on conscious experience, thought and action. The phenomenon of implicit memory is described and used as a basis to develop an analogous concept of implicit perception. In both cases the person shows the effects of current or past events, even though these events are not accessible to phenomenal awareness. There is collateral evidence for the emotional unconscious: emotional states can serve as evidence of implicit perception or memory, and there is evidence of desynchrony between the subjective experience of emotion, which can be identified with consciousness, and the effects of emotional responses on physiology and overt behaviour. Theoretical approaches to the psychological unconscious include a connectionist approach, which affords a limited role for conscious processing in mental life; a neuropsychological aproach, involving the disconnection of a module serving consciousness from the rest of the cognitive system; and a psychological approach, which emphasizes the central role in conscious awareness of mental representations of the self as the agent or experiencer of events.

1993 Experimental and theoretical studies of consciousness. Wiley, Chichester (Ciba Foundation Symposium 174) p 147-167

The doctrine of mentalism holds that mental states stand in relation to action as do causes to effects. That is to say, as Kant noted in his *Critique of Pure Reason*, that what we do is determined by knowledge, feeling and desire. Throughout most of the history of scientific psychology, most analyses of mental life have focused on conscious mental states. On the other hand, there has always been a substantial body of clinical and experimental evidence that our experiences, thoughts and actions are affected by mental structures and processes that lie outside of phenomenal awareness and voluntary control.

The cognitive and emotional unconscious

We may take a familiar example from cognitive neuropsychology: patients suffering bilateral damage to the medial temporal lobes, including the hippocampus,

148 Kihlstrom

show a gross anterograde amnesia, meaning that they are unable to remember even very recent events. Nevertheless, careful examination shows that they have preserved some sort of memory for these events. Thus, if they have studied a list of words incuding the item *ELASTIC*, they will fail to recall or recognize these items after a short period of distraction. However, if they are presented with three-letter stems, like ELA---, and asked to complete them with the first word that comes to mind, they are more likely to produce ELASTIC than, for example, ELATED. This advantage for studied over non-studied items is known as a priming effect. The typical outcome of such studies is that the magnitude of priming is approximately the same in amnesic patients as it is in intact subjects who show normal memory for the study list (for reviews, see Schacter 1987, Shimamura 1986). Interesting priming effects are preserved in the functional amnesias of dissociative disorders, as well as in the organic brain syndromes (Kihlstrom et al 1992a, Schacter & Kihlstrom 1989). In my laboratory, similar findings have been obtained in subjects with post-hypnotic amnesia (Kihlstrom 1980) and in surgical patients who have undergone general anaesthesia (Kihlstrom et al 1990); interestingly, there is as yet no evidence for implicit memory for sleep learning (Wood et al 1992).

Such results illustrate the distinction between two expressions of episodic memory, explicit and implicit (Eich 1984, Jacoby & Dallas 1981, Schacter 1987, 1992). Explicit memory refers to the person's conscious recollection of some past event, as commonly reflected in recall, recognition, or any other task that refers to some time in the past and asks the subject to describe what he or she did, or experienced, at the time. By contrast, implicit memory refers to any effect on experience, thought or action—priming effects or savings in relearning or other indications of transfer—that is attributable to some past event; these tasks made no reference to any particular past event, but memory for such an event is implicit in the subject's enhanced performance. A wide variety of experiments indicate that explicit and implicit memory are dissociable in at least two senses: (1) implicit memory can be spared even when explicit memory is grossly impaired; and (2) many experimental manipulations have differential effects on explicit and implicit memory.

Recently, my students and I have introduced the concept of *implicit* perception, analogous to implicit memory (Kihlstrom et al 1992b). The obvious example is the phenomenon of 'subliminal' perception, where a degraded stimulus (i.e. too weak, exposed too briefly, or obscured by metacontrast or a mask) cannot be consciously detected, but nevertheless influences performance on some task. More dramatic examples are found in the case of 'blindsight' following damage to the striate cortex, in the functional blindness, deafness and tactile anaesthesia observed in the conversion disorders, and in analogous phenomena produced by hypnotic suggestion. Explicit perception refers to the person's conscious awareness of an event in the current (or immediately past) environment, as reflected in the individual's ability to detect signals, identify

objects or describe the form, colour, distance or movement of stimuli. *Implicit perception* refers to a change in experience, thought or action that is attributable to a current event, independent of the person's conscious awareness of that event; again, perception is implicit in task performance. Our hypothesis is that explicit and implicit perception are dissociable, in much the same way as explicit and implicit memory are.

The explicit/implicit distinction can be extended to other domains. For example, several investigators have been concerned with *implicit learning*, in which subjects acquire complex rules and categories through experience, without being able to articulate the rules or category definitions themselves (Reber 1989, Lewicki 1986). Similarly, it appears that subjects can distinguish between soluble and insoluble problems, without being aware of the solution itself—a phenomenon that may be relevant to the experience of intuition in creative problem-solving (Bowers 1984). Taken together, these studies exemplify the cognitive unconscious (Rozin 1976, Kihlstrom 1987), in which perception, memory, learning and thinking influence ongoing experience, thought and action outside of phenomenal awareness.

Similar considerations support the notion of an emotional unconscious (Kihlstrom et al, unpublished 1993). Freud held that our behaviour is influenced by unconscious affects and drives, but this phenomenon can be understood without all the encumbrances of psychoanalytic theory. For example, the multiple-systems theory of emotion holds that emotional responses have three components subjective, behavioural and physiological—that are only imperfectly coupled (Lang 1968). This leaves open the possibility of desynchrony (Rachman & Hodgson 1974), in which, for example, the subjective feeling of fear disappears, but the emotional and physiological correlates of fear persist. For example, Weinberger et al (1979) identified a group of individuals as 'repressors' (low scores on anxiety but high scores on social desirability) who deny distress but respond behaviourally and physiologically in a manner resembling highly anxious individuals. This particular form of desynchrony is an example of the dissociation between two expressions of emotion: explicit (reflecting the conscious awareness of an emotion, feeling state or mood) and implicit (referring to changes in experience, thought or behaviour that are attributable to an emotional response).

Another aspect of the emotional unconscious may be observed when a person experiences an emotional response, but is unaware of the event that precipitated that response (e.g. Johnson et al 1985, Kunst-Wilson & Zajonc 1980). In this case, emotion serves as an index of implicit perception or memory. This is what the English poet Thomas Brown had in mind when, freely translating Martial, he stated 'I do not love you, Dr. Fell, but why I cannot tell'. It is also what Breuer and Freud meant when they concluded in their *Studies of Hysteria* that 'hysterics suffer from reminiscences'. In their view, hysterical symptoms were implicit expressions of memory for past traumatic experiences—memories

150 Kihlstrom

that were denied conscious access by virtue of repression, but which return to consciousness in disguised form.

Unconscious, preconscious, subconscious

One of the attractions of psychoanalytic theory was that it afforded a way of conceptualizing the nature of unconscious mental life. But now, 100 years later, other sorts of models are available. Traditional information-processing theory has generally accepted a distinction between automatic and controlled cognitive processes (e.g. Logan 1988, Shiffrin 1988). Automatic processes are executed inevitably under appropriate stimulus conditions, without requiring any intent or deliberation on the part of the subject. For example, individuals can decode written or spoken language without trying to do so; in fact, they cannot consciously prevent such a thing from happening. In the present context, the most interesting property of automatic processes is that we can have no conscious access to their operations. To continue the example, even very young children speak and listen fluently without consciously referring to the rules of grammar.

Alternative theoretical approaches to cognition also make room for unconscious processes. It has been proposed that some aspects of information processing are modular in nature (Fodor 1983, Jackendoff 1987). In theory, cognitive modules are isolated from other aspects of mental function: other modules have access to their products, but not to their internal processes. The same restriction applies to conscious awareness. A rather different perspective has been offered by parallel distributed processing approaches to cognition (e.g. McClelland & Rumelhart 1981, Rumelhart et al 1986), in which information is represented by a stable pattern of activation across a large number of task-specific processing units. By virtue of massive parallelism, both the number of processing units and the speed with which they operate exceed the span of conscious awareness. Only when the system relaxes to steady state, or when processing is slowed by virtue of stimulus ambiguity, does the representation become accessible to phenomenal awareness.

Automaticity, modularity and connectionism are based on rather different assumptions about the mind, but all assume that there is a vast repertoire of skills, rules, strategies and other processes that guide experience, thought and action outside of conscious awareness. The operation of procedural knowledge (Anderson 1983, Winograd 1975) may be described as *unconscious processing* in the strict sense of the term, because it is denied to introspective phenomenal awareness in principle, under any circumstances—it cannot be known directly, but only by inference. At the same time, it should be clear that the psychological unconscious should not be restricted to procedural knowledge. In implicit perception, it is the stimulus itself, not just the processes that analyse it, that is denied to conscious awareness; in implicit memory, it is the event itself, not just the processes that encode and retrieve it, that is denied to conscious

recollection. Obviously, some other categories are needed to represent completely the scope of unconscious processing.

Following Freud, we can identify as cases of preconscious processing those instances (e.g. subliminal perception, the amnesic syndrome) where the mental representation of an object or event has been degraded by experimental or subject conditions. For example, a stimulus may be presented at an intensity that is too weak, or for a duration that is too short, to permit conscious perception; or brain damage or experimental conditions may so impoverish encoding that the memory trace is inaccessible to conscious retrieval. But there are other instances, as in hypnosis or the dissociative and conversion disorders, where the event is in no sense degraded, but the event is denied conscious representation none the less. Following William James and Morton Prince, we can refer to these cases in terms of subconscious processing. In either case, it is clear that unconscious procedural knowledge can act on declarative knowledge that is not itself accessible to phenomenal awareness.

The distinction between preconscious and subconscious processing is important, because it addresses the question of restrictions on non-conscious processing. In many examples of implicit perception and memory, the extent of unconscious influence is severely limited. Greenwald (1992) (see also Kihlstrom et al 1992c) has concluded that there is good evidence for the preconscious processing of physical features of stimuli, and some evidence for simple semantic analyses, but no convincing evidence for the processing of complex meanings. Similarly, implicit memory in the organic amnesic syndrome appears to be largely supported by a perceptual representation system that retains information about the physical properties, but not the semantic features, of events (Schacter 1992, Tulving & Schacter 1990). Unconscious processing is not always analytically limited: for example, the priming observed in post-hypnotic amnesia (Kihlstrom 1980) clearly relies on the processing of semantic relationships.

Most likely, the scope of non-conscious processing depends on the precise manner in which the mental representations in question are rendered inaccessible to conscious awareness. When the stimulus is degraded (as in subliminal perception) or brain damage prevents optimal encoding (as in the organic amnesic syndrome), preconscious processing may be restricted to simple analyses, of the sort that can be performed by automatic or modular processes. We would not expect much beyond perception-based repetition priming during surgical anaesthesia, for example. Of course, to the extent that complex mental processes can become automatized through practice, the scope of preconscious processing may be expandable to some degree. Thus, if subjects have allowed certain complex judgements to become routine, these may be performed even on subliminal stimuli—especially if they are presented near the subjective rather than the objective threshold. But when conscious awareness is disrupted by a dissociative process, as in hypnosis, or the dissociative

152 Kihlstrom

and conversion disorders, rather more complicated forms of analysis appear to be possible.

Consciousness and the self

Although many issues remain to be resolved, experimental and clinical demonstrations of unconscious influence are now common. Theoretical developments have lagged somewhat farther behind. Traditional information-processing theories identify the unconscious with preattentive or automatic processing, or states of partial activation. Parallel distributed processing models seem to suggest that virtually all information processing is unconscious. Schacter (1990) has proposed a neuropsychological model, in which a cognitive module mediating conscious awareness can be disconnected from modules mediating perception, memory and the like. However, as noted earlier, not all instances of unconscious processing involve degraded stimuli or automatic, preconscious processing, and there are disruptions of consciousness that do not involve brain damage.

A psychological model that stays close to phenomenal experience suggests that consciousness is mediated by a particular knowledge structure, the *self* or the person's mental representation of him- or herself (Kihlstrom 1992). According to this view, the 'self structure' resides in working memory, where it routinely makes contact with mental representations of the local and global environment, the individual's current processing goals, and other knowledge structures activated by perception, memory and thought. This connection to the self, which identifies the self as the agent or experiencer of some represented event, appears to be the key to consciousness. After all, as William James noted, consciousness is a consciousness of *something*, but it is also *personal*: consciousness comes when we take possession of our thoughts, feelings and desires, and act-nowledge them as our own.

Interestingly, both Ernst Claparede (1950) (discussing Korsakoff's syndrome) and Pierre Janet (1907) (in his treatises on hysteria) noted the absence of the self in cases of unconscious influence. Following their lead, we may propose that when a link is made between the mental representation of self and the mental representation of some object or event, then the percept, memory or thought enters into consciousness; when this link fails to be made, it does not. Nevertheless, unconscious percepts and memories, images, feelings and the like can still influence ongoing experience, thought and action—perhaps by serving as sources of activation which spreads to other knowledge structures and activates them.

The situation can be portrayed in an associative-network model of memory operating according to a principle of spreading activation, much like ACT* (Anderson 1983). Knowledge is represented by nodes standing for concepts, and by associative links for the relations between them. Fig. 1 shows the three basic

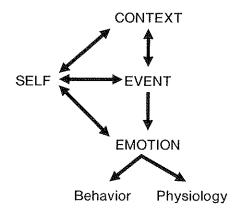


FIG. 1. Schematic representation of an event linked to co-activated mental representations of self and episodic context, and giving rise to an emotional state.

elements in the mental representation of a particular event (whether constructed in perception or reconstructed from memory): an event node, representing a raw description of the event; a context node, representing the spatiotemporal context in which the event occurred; and the self node, representing the self as the agent or experiencer of the event. These nodes are each linked to many other nodes in the memory network, to which activation can spread. For example, if the event concerns a doctor, activation may spread to associated concepts like nurse or hospital; or the connotative meaning of the event may give rise to an emotional state, with behavioural and physiological consequences.

Now, imagine that for some reason the link to the self does not get made. Perhaps the stimulus is so degraded that it does not get represented in working memory or perhaps the person is anaesthetized for surgery. Alternatively, the link may get made at the time of perception, but it may not be stored in permanent memory: hippocampal damage may prevent the consolidation of such an association, or the formation of the association may require a certain degree of elaborative processing at the time of encoding. Or, the link may be stored in permanent memory, but inaccessible at the time of retrieval. Perhaps the spread of activation down the associative pathway is inhibited by some sort of dissociative process; or, as may occur in cases of genuine multiple personality. there may be two or more self-concepts simultaneously resident in memory, but the one that is active at the moment is not linked to the knowledge in question. In any of these cases, the person is not going to be conscious of the event in question. Nevertheless, the event node may serve as a source of activation that spreads to other nodes, producing the phenomena of priming that exemplify the cognitive unconscious. Or, the emotional state generated by the event may have behavioural and somatic effects, producing the phenomena of desynchrony that exemplify the emotional unconscious.

154 Kihlstrom

The general framework set out here certainly requires elaboration and experimental testing; developments would be facilitated by implementation in an operating computer simulation. That is on the agenda for the future. For the present, we may be content with three conclusions. First: there is now plenty of clinical and experimental evidence for the influence of unconscious perception and memory on the person's experience, thought and action. This evidence is not confined to a single paradigm, but is derived from a number of different methods, ranging from subliminal perception to hypnosis, from neuropsychology to multiple personality, each shedding unique light on the problem. Second: the psychological unconscious, as we now understand it, is quite different from the dynamic unconscious envisioned by classical psychoanalysis. We can honour Freud's insights about unconscious influences, at the same time as we free ourselves from the confines of his theory. Third: the operations of the psychological unconscious are not curious anomalies, but are central features of mental life that can be accommodated within the framework of contemporary cognitive theory. The role of the self in conscious experience provides a basis for uniting the two cultures of modern psychology: cognitive neuropsychology on the one hand, and personality and social psychology on the other.

Acknowledgements

I thank Terrence Barnhardt, Lawrence Couture, Jennifer Dorfman, Martha Glisky, Elizabeth Glisky, Lori Marchese, Sheila Mulvaney, Robin Pennington, Michael Polster, Victor Shames, Susan Valdesseri and Michael Valdesseri for their comments. The point of view represented here is based on research supported by grant #MH-35856 from the National Institute of Mental Health.

References

- Anderson JR 1983 The architecture of cognition. Harvard University Press, Cambridge, MA
- Bowers KS 1984 On being unconsciously influenced and informed. In: Bowers KS, Meichenbaum D (eds) The unconscious reconsidered. Wiley, New York, p 227-272
- Claparede E 1950 Recognition and me-ness. In: Rapaport D (ed) Organization and pathology of thought. Columbia University Press, New York, p 58-75 (Original work published 1911)
- Eich E 1984 Memory for unattended events: remembering with and without awareness. Mem Cognit 12:105-111
- Fodor J 1983 The modularity of mind: an essay on faculty psychology. MIT Press, Cambridge, MA
- Greenwald AG 1992 New look 3: a paradigm shift reclaims the unconscious. Am Psychol 47:766-779
- Jackendoff R 1987 Consciousness and the computational mind. MIT Press, Cambridge, MA
- Jacoby LL, Dallas M 1981 On the relationship between autobiographical memory and perceptual learning. J Exp Psychol Gen 110:306-340

- Janet P 1907 The major symptoms of hysteria. Macmillan, New York
- Johnson MK, Kim JK, Risse G 1985 Do alcoholic Korsakoff's syndrome patients acquire affective reactions? J Exp Psychol Learn Mem Cognit 11:22-36
- Kihlstrom JF 1980 Posthypnotic amnesia for recently learned material: interactions with 'episodic' and 'semantic' memory. Cognit Psychol 12:227-251
- Kihlstrom JF 1987 The cognitive unconscious. Science (Wash DC) 237:1445-1452
- Kihlstrom JF, Klein SB 1992 The self as a knowledge structure. In: Wyer RS, Srull TK (eds) Handbook of social cognition, 2nd edn. Erlbaum, Hillsdale, NJ, in press
- Kihlstrom JF, Schacter DL, Cork RC, Hurt CA, Behr SE 1990 Implicit and explicit memory following surgical anesthesia. Psychol Sci 1:303-306
- Kihlstrom JF, Tataryn DJ, Hoyt IP 1992a Dissociative disorders. In: Sutker PB, Adams HE (eds) Comprehensive handbook of psychopathology, 2nd edn. Plenum Press, New York, p 203-234
- Kihlstrom JF, Barnhardt TR, Tataryn DJ 1992b Implicit perception. In: Bornstein R, Pittman TS (eds) Perception without awareness. Guilford Press, New York, p 17-54
- Kihlstrom JF, Barnhardt TR, Tataryn DJ 1992c The psychological unconscious: found, lost, and regained. Am Psychol 47:788-791
- Kunst-Wilson WR, Zajonc RB 1980 Affective discrimination of stimuli that cannot be recognized. Science (Wash DC) 207:557-558
- Lang PJ 1968 Fear reduction and fear behavior: problems in treating a construct. In: Shilen JM (ed) Research in psychotherapy. Am Psychol 3:90-102
- Lewicki P 1986 Nonconscious social information processing. Academic Press, New York
- Logan GD 1988 Towards an instance theory of automatization. Psychol Rev 95:492-527
- McClelland JL, Rumelhart DE 1981 An interactive activation model of context effects in letter perception. I. An account of basic findings. Psychol Rev 88:375-407
- Rachman S, Hodgson SI 1974 Synchrony and desynchrony in fear and avoidance. Behav Res Ther 12:311-318
- Reber AS 1989 Implicit learning and tacit knowledge. J Exp Psychol Gen 118: 219-235
- Rozin P 1976 The evolution of intelligence and access to the cognitive unconscious. In: Sprague JM, Epstein AN (eds) Prog Psychobiol Physiol Psychol 6:245-280
- Rumelhart DE, McClelland JL, and the PDP Research Group 1986 Parallel distributed processing—explorations in the microstructures of cognition, vol 1: Foundations. MIT Press, Cambridge, MA
- Schacter DL 1987 Implicit memory: history and current status. J Exp Psychol Learn Mem Cognit 13:501-518
- Schacter DL 1990 Toward a cognitive neuropsychology of awareness: implicit knowledge and anosognosia. J Clin Exp Neuropsychol 12:155–178
- Schacter DL 1992 Understanding implicit memory: a cognitive neuroscience approach. Am Psychol 47:559-569
- Schacter DL, Kihlstrom JF 1989 Functional amnesia. In: Boller F, Grafman J (eds) Handbook of neuropsychology. Elsevier Science Publishers, Amsterdam, vol 3: 209-231
- Shiffrin RM 1988 Attention. In: Atkinson RC, Herrnstein RJ, Lindzey G, Luce RD (eds) Stevens' handbook of experimental psychology, vol 2: Learning and cognition, 2nd edn. Wiley, New York, p 739–812
- Shimamura AP 1986 Priming effects in amnesia: evidence for a dissociable memory function. Q J Exp Psychol A Hum Exp Psychol 38:619-644

Tulving E, Schacter DL 1990 Priming and human memory systems. Science (Wash DC) 247:301-305

Weinberger DA, Schwartz GE, Davidson RJ 1979 Low-anxious, high-anxious, and repressive coping styles: psychometric patterns and behavioral and physiological responses to stress. J Abnorm Psychol 88:369–380

Winograd T 1975 Frame representations and the declarative-procedural controversy. In: Bobrow D, Collins A (eds) Representations and understanding. Academic Press, New York, p 185-210

Wood JM, Bootzin RR, Kihlstrom JF, Schacter DL 1992 Implicit and explicit memory for verbal information presented during sleep. Psychol Sci 3:236-239

DISCUSSION

Searle: In what I take to be the main thesis of your paper, the claim is that there are these unconscious mental events. The emphasis has to be on 'mental'. I want to describe two different hypotheses; you tell me why you should prefer one rather than the other.

One hypothesis is: there are all sorts of things that have an impact on the nervous system, and leave effects on the nervous system, that don't come into consciousness. Nonetheless, they affect subsequent conscious thoughts, feelings and behaviour. That's the hypothesis I think is right. For example, when I am sound asleep, my memories exist in my brain in forms that are capable of causing conscious memory experiences, even though there is nothing mental occurring at that time while I am asleep.

The second hypothesis is that in addition to the level of neurophysiological description and all of the effects on it and of it, there is an unconscious mental level. Freud actually said that all mental states are in themselves unconscious and that what we call consciousness is like perception. You go into the attic where all your unconscious things are and you shine the light of perception on these things—that's what bringing a mental state to consciousness is.

I thought all the evidence you gave supported the first of those views. Then at the very end it seemed to me you blew a trumpet in favour of the second.

Kihlstrom: I certainly wouldn't want to be remembered as having blown a trumpet in favour of anything psychoanalytic.

I do want to insist that these things are mental in a sense that they are representations of experience. They are constructed through perception or reconstructed through memory or whatever, and simply have an impact on the person's ongoing experience without themselves being conscious. I don't think this is the same as playing with somebody's neurotransmitter system, which would not be mental in that sense. These are representations. I think the best evidence for the representational status is the fact that you can get conscious access to them under some conditions. The whole beauty of hypnosis and hysteria, and the reason I'm interested in hypnosis and hysteria, is that we can

can be recalled, they must have been conscious. Their non-consciousness is purely a split in consciousness.

Kihlstrom: There is an important vocabulary problem here. This is one William James talks about in the *Principles*: he offers 10 disproofs of the idea that there is an unconscious. Then in the very next paragraph, he says, 'nevertheless, there is this thing we have to give a name to, maybe it's coconscious, maybe it's sub-conscious'. When I use the term unconscious, I'm using it in the broad, historically valid, psychological sense of not accessible to conscious awareness.

Williams: Can I make a distinction, which we ought to be able to make at this point in the proceedings? There is one question, which you and John Searle are discussing—the question of whether there are psychological or mental processes which are not consciously registered. Personally, I agree with you, and disagree with Searle, in thinking that there are such processes. This seems to be a separate issue from the existence of something called 'the unconscious', which is invoked in psychoanalytical theory.

Kihlstrom: I never used that term, except to define a domain of inquiry. Williams: One reason for separating the questions is that in Freud's doctrine about the unconscious, there is a necessary conceptual link between the idea of the unconscious and the idea of repression: things are in the unconscious for reasons that are connected with motivation of various kinds. Whether we should think of the mind in such terms seems to me a separate question from the one that's under discussion in your paper.

Kihlstrom: I view unconscious as an adjective, not a noun. Unconscious and conscious are qualities of mental life.

Shevrin: I would like to address John Searle's question because I think it's critical, at least from my point of view as a psychoanalyst, and because the question bears on my empirical efforts to explore this question. Is the unconscious simply to be considered dispositional, such that we know about it only by its influence on consciousness? Or is the unconscious to be considered another 'realm' of unconscious mental life, having the ability, for example, to influence complex and meaningful actions in a way that bypasses consciousness? One could cite clinical evidence that actions (behaviours in the strict sense) can be influenced by events, which the person is not aware of at the time but which are active at the time, and not simply dispositional. Compulsions, phobias, enactments of post-hypnotic suggestion and sleep-walking are examples.

There is an analogy that captures the difference, and I think the difference is important. I went to the theatre yesterday: the stage was set and the actors were on stage. There is a great deal of unseen stage machinery which makes that possible. The stage machinery is the dispositional elements, what you see lit up in front of you is consciousness. In the Freudian notion of the unconscious, there are two stages, or there can be three or four stages: there is the lit stage and, alongside, there is another stage that is in darkness, on which things are

cancel the amnesia suggestion and the person remembers the word list perfectly well. Similarly, when a person recovers from a fugue state, the person remembers his or her past perfectly well. You have to give those things the status of mental representations; they are not merely neurotransmitters hanging around in the brain matter.

Searle: If you ask what fact about the actual state, at the time that it's unconscious, makes it mental, the answer you gave was (I think the correct answer) a dispositional answer. That is, there are certain circumstances in which you are capable of bringing that state to consciousness. Take a very unproblematic case: I believe, even when sound asleep, that George Bush is President of the United States. There is no little sentence in my head that I'm thinking quietly about; the only fact about me when I'm asleep is that I have neurophysiological structures that are capable of bringing that belief to consciousness. Those are two different views and they are not equivalent. The evidence supports the weaker view, namely, attributions of unconscious states are attributions of a dispositional capacity of neurophysiological structures to cause conscious states. So the opposition that you make between a genuine mental state and these neurophysiological phenomena is a false opposition. There are neurophysiological structures in the brain capable of producing conscious thoughts and behaviour. It is perfectly reasonable to call those structures unconscious, even unconscious thoughts and beliefs. The mistake is to suppose that somehow the evidence you have given supports a Freudian style

Kihlstrom: Never, in my entire career, have I said that this evidence supports a Freudian view!

Searle: Then leave out Freud! On your view, is there supposed to be an ontological realm of unconscious mental states that have all their mental aspects in addition to the neurology?

Kihlstrom: One of the reasons I was interested in doing those experiments was that it brought to light an interesting divergence between how some philosophers—not necessarily you— and some psychologists define things as mental. I think it's a very big mistake to identify the mental with the conscious; it is psychologically not correct. Here you have percepts and memories and emotions that have, by any behavioural standards, psychological reality, even though the person is not consciously aware of them. That is a very striking empirical fact from psychology that people ought to pay attention to.

Marcel: John, if you don't support Freud, I have heard you support Pierre Janet. It is not that these priming effects are unconscious or mental, but that there is a split consciousness. The mere fact that x facilitates a subsequent y does not mean that x is mental. This is the case for priming. It's like the fact that when I do press-ups, I get better at doing press-ups. There is nothing mental about that. I want to know what criteria you have that these priming effects are mental. Secondly, the evidence you use to show the priming effects are mental is that people can perfectly well recall them under certain circumstances. If they

happening, interacting with and influencing what is occurring on the lit stage. Actors from the lit stage disappear onto the unlit stage and what goes on there influences the action on the lit stage, although we can only infer from events on the lit stage what might be happening on the unlit stage.

I would like to cite some empirical evidence to support this Freudian view of the unconscious. With evoked potentials, you can detect the activity of the brain when an individual is not conscious of a subliminal stimulus. When certain words of particular importance to an individual's unconscious conflict, as determined by psychoanalysts, are presented tachistoscopically, the brain responds differently when they are subliminal to when they are supraliminal. A discriminant analysis revealed that the brain responses to the words related to the unconscious conflict are better categorized as going together when they are subliminal than when they are presented supraliminally. The reverse relationship is found for brain responses to words selected to reflect the patient's conscious experience of a particular symptom believed to be caused by the unconscious conflict (Shevrin et al 1992). What is instantiated by the evoked potentials is an act of categorization which is psychological in nature. The process is ongoing and active; it isn't simply dispositional, nor is it in any way reflected in consciousness. I would like to submit that as evidence in favour of the Freudian unconscious as a set of mental operations rather than as a set of dispositions.

Dennett: John Searle talks about events being mental if they have the disposition to become conscious, if the person is in a state with content such that that very content can become conscious. He overlooks another possibility: the feature of these events that I would cite as a basis for calling them mental is that they are content specific. They are the result of a categorization process; there is a difference between the state of 'ocean' and the state of 'tower', if those are your primes. The fact that one of them is 'ocean' is not necessarily tied to the fact that that's the content of that unconscious brain process. The content may not be able to be made conscious, but it may be able to influence something else that becomes conscious, by being semantically related to it.

Searle: That's a dispositional analysis.

Lockwood: John, if one asks: 'Is this an unconscious mental state or is it just a neurophysiological state that has the power to produce certain actions and conscious states under appropriate circumstances?', it's easy to persuade oneself that there isn't a real issue here at all. But it seems to me there may be a real issue, if one looks instead at the dynamics of what is being proposed here. Suppose that our best psychological theory says that there are states which are not directly accessible to consciousness but which give rise to behaviour and conscious thoughts and feelings, and evolve over time, in ways that can be most naturally and straight-forwardly explained by using essentially the same (folk-psychological) language and categories that we customarily employ at the conscious level. Then, it seems to me that one is thereby giving substantive

content to the supposition that these states are, in some sense, genuinely mental. (Of course, speaking as a materialist, this doesn't prevent them being neurophysiological as well.)

Marcel: But surely the data go against the notion that they are semantic. If you look at the literature on priming effects, people like John Williams have shown you don't get unconscious semantic priming effects. There are associative priming effects, but they are not semantic, and that's a crucial distinction.

Gray: For the purpose of this argument, associative priming can do exactly what Dan was trying to do with semantic priming.

Marcel: No. It's not representation. This depends on your criterion for the mental. I understood John Kihlstrom to say the criterion was one of representation. If the priming is purely associative, then I don't see why there is anything necessarily representational about it. And if so, it's not mental.

Dennett: How do you distinguish pure association from the semantic priming effects?

Marcel: One examines whether there is a difference in effectiveness of priming between stimuli that are associated but do not co-refer and stimuli that co-refer but are not associates. It appears that you get automatic priming for associates when the prime is unconscious, but not for pairs of stimuli linked only by their meaning.

Humphrey: I want to get back to John Kihlstrom's interesting remarks about the self and its relation to unconscious processes. His ideas are quite well supported by some of the data from patients with blindsight, particularly the evidence of the difference between human beings and animals. I worked many years ago with monkeys who had had the striate cortex removed: they retained extraordinarily sophisticated visual capacity, much better than anything which has yet been discovered in human beings with lesions of the striate cortex. One way of thinking about this is that the monkey has an advantage in that it doesn't have a particularly highly developed concept of self. Hence, the monkey's nonsensory visual percepts are nothing like so surprising to the monkey as to the human. For a human to have a percept which isn't his own percept (related to himself) is very odd indeed. So human patients retreat into saying, 'I don't know what's going on' and denying their ability to see at all. For the monkey, I suspect perceptual information doesn't create the same sort of existential paradox, therefore the monkey is much more ready to use it.

Interestingly, for one particular monkey I worked with for a long time, there were conditions under which she became unable to see again—if she was frightened or she was in pain. It was as though anything which drew attention to her self undermined her ability to use unconscious percepts.

Dennett: For people who haven't seen the film of this monkey, Helen, it is important to stress how amazing she is. Nick and I showed a film of this monkey to a group of primate experts. We said when we started the film: 'There is something wrong with this monkey, what is it?' Nobody guessed that she was

blind; her behaviour was absolutely consistent with her having perfectly good vision. The only thing they noticed was when there was a bit of sticky tape on the floor she kept trying to pick it up, and failing.

Gray: Dan, doesn't your allegiance to the Turing test therefore make you say she isn't blind?

Dennett: Manifestly, she isn't blind.

Kihlstrom: The reason she's not manifestly blind is that her circumstances are such that she's able to capitalize fully on implicit perception.

Marcel: Nobody has ever asked a cortically blind monkey whether it can see, though Alan Cowey has proposed a way of doing it.

There are two different points to be made. First, in all human cases studied, the patients are blind only within a circumscribed area, maximally a half-field. So they can see in their preserved field. Second, neither do they have no experience of stimuli in their blind field, nor do they have percepts that are not experienced as their own. They often experience stimuli presented to the blind field, but the experiences are not 'visual'; the patients say they are 'felt'.

Harnad: There is a good example in spatial location. I have read that patients with blindsight can localize objects (say, on the left or right) because they feel some sort of inclination or urge to orient towards them.

Marcel: Sorry, that is not the case. There has never been a problem with location. No blindsight patient has ever been described who had a scotoma that went over the body midline.

Harnad: It doesn't have to be over the midline; their sensation is an inclination to turn in the direction of the object.

Marcel: Turning in the direction is not the issue; it's often a question of what shape the stimulus is.

Shevrin: The interesting findings about implicit perception and implicit memory that John Kihlstrom described raise a question in my mind. In a psychological sense, why does it work? Why are we able to retrieve information in consciousness in one way and not in the other? If the information is available, why does it take some kind of indirect means, like free associations or word stems, for the unconscious to be retrieved?

Kihlstrom: I worry a little about the use of the word 'retrieval'. I'm not sure that information is ever retrieved in the usual sense. Consider, for example, an amnesic patient who hears the word 'water' and then produces the word water as opposed to 'watch' on a stem-completion test. That patient never retrieves the memory in the same way that an analytic patient is supposed to retrieve a repressed childhood recollection. They never say, 'Oh yes, I remember, it was water'.

Shevrin: That's not what we rely on psychoanalytically: that happens in very rare instances. As I have described (Shevrin 1992), it is necessary to distinguish the content, or information, from the vehicle through which the content becomes conscious. The vehicle may be a perception, thought, memory, image, etc. What

most often happens in psychoanalysis is that a particular content, belonging to a childhood event, will emerge as an image or in a dream, not necessarily as an explicit memory. There are many examples in psychoanalytical work in which a patient will forget in one session entirely about some events described with much feeling in the preceding session, yet the psychoanalyst will detect clear evidence of the content in the patient's various associations. In short, the analyst is relying on the patient's implicit memories.

Kihlstrom: And interpreting them in some ways in the same way as the experimenter does. This idea of retrieval has a kind of intentionality to it that just seems wrong for what seems to be a passive operation of an associative network.

Shevrin: So what's your explanation?

Kihlstrom: I think the simple-minded answer is, implicit perception works by spreading activation.

Shevrin: But why isn't spreading activation helpful in explicit consciousness? Kihlstrom: We have talked about first-person and third-person analyses of consciousness, we have forgotten the second person. Suppose I asked you, 'Howard, what are you aware of right now?'. In an associative network theory of memory all this information is supposed to get unpacked. The query: what do you see now? is supposed to light up little nodes in the associative network. One of them corresponds to 'you', which is the self node; others correspond to the things you are seeing right now. But if there is no connection, if there is no link between the self node and the node that's stimulated by what you are seeing, you can't answer that question. That's how you programme it in Anderson's ACT* computer simulation (Anderson 1983). You create a self-structure to which all percepts and memories running through working memory will be related. If the association is preserved, you are able to report; if it is not preserved, you are not.

Gray: The issue of what are non-conscious mental operations is very important. There is at least one way—I think the right way—in which we haven't yet talked about it. We are basically concerned with the scientific and conceptual issues of what is consciousness, what is the mental and so on. Among those issues is the question, what level of analysis should we work at? John Searle's point, certainly in his writings as I have understood them, is that you need only two levels of analysis: the level of analysis of the neurophysiologist, neuroanatomist, neurochemist and so on, and the level of fully conscious mental states. I have disputed that (Gray 1987) and I would like to summarize the argument in relation to something which is nicely neutral with respect to all of these issues about Freud and psychoanalysis (which can obscure the main focus of what I take this conference to be about).

The dispute in print has been about language processes. John Searle says there are brain events going on and out comes language, but it is a category mistake to talk about the brain as doing phonetics, syntax and semantics. These are

clearly mental operations—nobody could deny that syntax and semantics are something different than simply neurophysiological processes. So the dispute has been: is it right to treat the brain processes that lead to spoken language as themselves having a level of analysis that is beyond the neurophysiological? My argument is as follows.

Clearly, anything that is going on in the brain to produce syntax or semantics is part of neurophysiology; there are nerve cells doing all the things that nerve cells have to do. Then, however, one must ask: why are the nerve cells doing those things and not others? The answer to that question is not in terms of neurophysiology, it is in terms of constraints that are required for communication between individuals, because that's what speech is about. The constraints on speech between individuals include a level of syntax, because without syntax you don't have the informational combinatorial capacities that you need for language, and a level of semantics, because without that you don't have shared referents. Neither the semantic nor the syntactic properties that are necessary are properties of neurophysiological events; they are properties of the communication system. Therefore to analyse them you need the level of analysis of mental operations. In that sense, there is unconscious mental processing—syntactical and semantic properties are mental, clearly, and they are unconscious because we haven't the slightest idea of how the syntax or semantics is working as we talk.

Searle: I didn't actually disagree with anything you just said. Maybe I'm not making myself clear or maybe I haven't understood what you said. My view is as follows. It's clear that we will have to have different levels of description of what is going on in the brain, and some of those levels of description are at a much higher level than that of standard neurophysiological text books. But, at the same time, some of the higher levels are not at the level of consciousness. For example, in order to describe the brain's unconscious capacity to form relative clauses, we need to talk at a much higher level than the level of neuron firings. There has to be a much higher level of description than that.

I am asking what facts correspond to these claims. What's the mode of existence of the phenomena that we are describing here? I have said the only facts that I have ever seen anybody be able to identify are facts involving neurophysiology, described at these different levels, and facts about consciousness. I can't find any other mental facts. This point comes out especially strongly when we examine rule-governed behaviour. It's clear that we do follow rules; it's clear that sometimes we follow rules unconsciously. However, since Chomsky, it has become fashionable to attribute rules and rule following behaviour to people, where the rules aren't even the kind of thing that could ever be the content of a conscious mental state. On Chomsky's view, it isn't that the child doesn't happen to think 'move alpha'; it couldn't think 'move alpha', because the expression 'move alpha' is just the linguists' way of describing the rule. The rule itself isn't the kind of thing that you could ever

follow consciously. Against this view, I want to say that it is not a rule, rather, there is a pattern of behaviour that is produced by neurophysiological structures, where the structures have neurophysiological but not mental reality.

Kihlstrom: On those grounds, I think the kinds of things I'm talking about here satisfy your criterion of the mental. If the stimulus were not masked or were a little stronger or the person didn't have so much hippocampal damage, what had been presented would be perfectly perceptible or perfectly memorable.

Harnad: It seems to me that if (according to the best current empirical theory in linguistics) 'move alpha' is a useful, predictive, explanatory descriptor of a process going on in one's head and playing a role in generating one's linguistic capacity, then it surely does not matter whether it is 'mental' (i.e. conscious or potentially conscious) or not. 'Move alpha' could be a correct 'higher-level description of neural activity' either way.

Searle: It matters to Chomsky.

Williams: The point was made earlier that consciousness seems to be either all or nothing. The classical argument for unconscious perceptions, which was given by Leibniz, goes against that claim. You can be woken up by a bell. Leibniz's view was that if you were woken up by a bell, you were woken up by hearing a bell and you gradually became conscious of hearing a bell. This seems an absolutely commonsensical description of the situation; it has three consequences. It implies that there is such a thing as hearing when you are not conscious that you are hearing. Second, it implies there are degrees of consciousness. Third, it implies there can be an unconscious mental activity which is not analysable into disposition.

Marcel: You are right. The evidence psychologically is against what Ben Libet said. He said, the production of a conscious experience has a unitary, all or nothing character. That is not the case.

Singer: I think if there is a thrust to the work that I have been doing, it is that people do much more preparation at a conscious level for all kinds of subsequent acts than they realize. We overlearn all kinds of mental associations, social connections, potential situations and so on, in much the same way that we overlearn the movements required to ride a bicycle. A great deal of the time we are preparing ourselves for sets of, for example, speech patterns. While it can be said that we don't pay attention to our speech, that it seems to come out without thought, this may be because the material to be spoken was often rehearsed extensively beforehand. I have often prepared, mentally, things I might say in certain circumstances. There is a tremendous amount of research on the advantages of mental practice. Imagery practice for atheletes, for example, shows that there are advantages, in addition to physical practice, of mental practice. This is not inconsistent with what John Kihlstrom has been saying. It could be argued that it helps us understand why certain things seem to operate automatically without subsequent specific recollection of their initial practice or rehearsal.

Secondly, I'm really not overwhelmed by the Leibnizian argument. What's the factual difference between saying, 'There is an unconscious hearing of the bell and eventually because the guy is hearing it unconsciously he wakes up' and saying 'There is a stimulation of the guy's nervous system which eventually reaches the threshold of consciousness and he wakes up'? I can't hear any factual difference.

Let me give you another example, closer to real life. Many of us suffer from back pains; sometimes the pain wakes you up. How should one describe that? It seems to me there are two different descriptions. One is this: 'I have an unconscious pain. It keeps going all night long and sometimes it gets so bad that it wakes me up'. A second description is this: 'I have this stimulation of my sciatic nerve, and eventually it reaches the threshold of consciousness, at which point I wake up'. I can't hear anything that's added (that isn't a source of confusion) by saying, in addition to the neurophysiology, there is this extra mysterious thing, the unconscious pain.

Van Gulick: I would like to tie together a few pieces of the discussion. Yesterday, our focus was on subjectivity. Today we have been focusing on a variety of issues to do with the conscious/unconscious distinction. John Searle has been arguing that there is no way to make sense of unconscious mental states, except as dispositions that give rise to conscious mental states. Dan Dennett, John Kihlstrom and some others have proposed that representational characteristics can be present in unconscious states that cannot be brought to consciousness. I think that's one of the strong arguments for the existence of non-dispositional unconscious states. One of the two traditional marks of the mental is the intentional or representational. There are many patterns that John Searle is willing to acknowledge at higher levels of neural organization that he wants to keep as non-mental, but it's very hard to see how you can possibly capture the relevant patterns without describing them in representational or intentional or semantic terms.

On the other hand, there's also the other traditional mark of the mental: the qualitative or subjective. There are reasons, like the one Sydney Shoemaker gave, for thinking that there are lots of qualitative states that may be mental in the qualitative sense but are unconscious in the sense that they cannot be referred to self and they cannot be reported. A good example is the work of Stoerig & Cowey (1990) on apparent chromatic sensitivity in patients with blindsight who report themselves as being blind and not having any experiences, and yet can show very good colour discrimination.

Marcel: They show wavelength discrimination, not colour discrimination. Gray: I have been feeling glimmerings of a possible consensus, or at least a hypothesis which would integrate a number of things that people have been saying from very different angles. Like many other things in this field, I suspect it goes back to something Larry Weiskrantz (1986) wrote.

You get into a lot of traps, if you think of consciousness as being primarily about stimulus detection and response execution. Max Velmans (1991) wrote a very good paper, in which he went systematically through all the things that consciousness is supposed to be about in terms of stimulus detection and response execution; he showed that consciousness was always too slow, it came too late. The obvious reply is to say consciousness is not to do with stimulus detection or response execution at all, because those things happen unconsciously. (And I would continue to say that they happen unconsciously through mental processes.) Therefore, what consciousness is doing is something to do with monitoring after the event. The key word probably is monitoring, and I think that's the word Larry Weiskrantz used—monitoring and predicting what's about to happen next; looking to see whether the world matches the prediction and detecting mismatch. All of those things take some time after the actual stimulus input. In the kind of experiment Ben Libet described (this volume), detection of the stimulus, i.e. the first bit of neuronal processing, goes all the way up to the sensory cortex. Consciousness comes in after that and is then used to generate further processes which are concerned with monitoring whether what has just come in is what ought to have come in and seeing what it predicts about what next ought to come in. All that, I take to be part of what's occurring in the Kinsbourne-Dennett view of multiple drafts (Dennett 1991, Dennett & Kinsbourne 1992). The multiple draft theory is about trying to get the best handle on what is going on and what ought to be going on next. So there is plenty of time for the editing, and some of that may be the time that was described in Ben Libet's cortical stimulus experiments. Some of these ideas I will try to develop in my paper tomorrow (Gray, this volume). I feel that if we stop thinking about consciousness as to do with stimulus detection and action, we might get somewhere.

Humphrey: Jeffrey, what has a stab of pain got to do with this kind of cognitive monitoring?

Gray: I was expecting that question, because it's precisely the one I have been asking myself for a number of years. I think the answer might be that a system which is concerned with predicting what ought to happen next and keeping, as it were, the whole ship steering along the right course, is a system which must also have a very important interrupt mechanism for times when what comes in is not according to plan. The pain itself we know occurs after the response—it is too late to organize the removal of the finger from the hotplate. So I think even pain can be fitted in; it occurs after the event, and it is telling you that the action pattern you have been carrying out is the wrong action pattern, you had better review that.

Marcel: There is no point in trying to stipulate what consciousness is. People here are referring often to different things, which may be related or may not be. Some people are referring to sensation, phenomenal aspects, some may be referring to intentionality, some to reflexivity, some to reportability. I would

prefer that we didn't use the word consciousness, but referred to which of those we are talking about. Otherwise, there is a systematic confusion and ambiguity.

Gray: Some of us still believe that consciousness matters, that the distinction between neurophysiological processes that don't have consciousness tacked on and those that do, matters.

Marcel: But which do you mean by the word consciousness?

Kihlstrom: I think it would be a disaster if we identified consciousness with reportability. People can be conscious of things they can't report. If you mean reportable in principle, perhaps we can agree.

References

Anderson JR 1983 The architecture of cognition. Harvard University Press, Cambridge, MA (Cognit Sci Ser 5)

Dennett DC 1991 Consciousness explained. Allen Lane, London

Dennett DC, Kinsbourne M 1992 Time and the observer: the where and when of consciousness in the brain. Behav Brain Sci 15:183-201

Gray JA 1987 The mind-brain identity theory as a scientific hypothesis: a second look. In: Blakemore C, Greenfield S (eds) Mindwaves. Blackwell Publishers, Oxford, p 461-483

Gray JA 1993 Consciousness, schizophrenia and scientific theory. In: Experimental and theoretical studies of consciousness. Wiley, Chichester (Ciba Found Symp 174) p 263-281

Libet B 1993 The neural time factor in conscious and unconscious events. In: Experimental and theoretical studies of consciousness. Wiley, Chichester (Ciba Found Symp 174) p 123-146

Shevrin H 1992 Subliminal perception, memory, and consciousness: cognitive and dynamic perspectives. In: Bornstein RF, Pittman TS (eds) Perception without awareness: cognitive, clinical and social perspectives. Guilford Press, New York, p 123-144

Shevrin H, Williams WJ, Marshall RE, Hertel RK, Bond JA, Brakel LA 1992 Event-related potential indicators of the dynamic unconscious. Consci & Cognit 1:340–366 Stoerig P, Cowey A 1990 Wavelength sensitivity in blindsight. Nature 342:916–918 Velmans M 1991 Is human information processing conscious? Behav Brain Sci 14:651–726 Weiskrantz L 1986 Blindsight: a case study and implications. Oxford University Press, Oxford



EXPERIMENTAL AND THEORETICAL STUDIES OF CONSCIOUSNESS

A Wiley-Interscience Publication

1993

JOHN WILEY & SONS

©Ciba Foundation 1993

Published in 1993 by John Wiley & Sons Ltd Baffins Lane, Chichester West Sussex PO19 1UD, England

All rights reserved.

No part of this book may be reproduced by any means, or transmitted, or translated into a machine language without the written permission of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, USA

Jacaranda Wiley Ltd, G.P.O. Box 859, Brisbane, Queensland 4001, Australia

John Wiley & Sons (Canada) Ltd, 22 Worcester Road, Rexdale, Ontario M9W 1L1, Canada

John Wiley & Sons (SEA) Pte Ltd, 37 Jalan Pemimpin #05-04, Block B, Union Industrial Building, Singapore 2057

Suggested series entry for library catalogues: Ciba Foundation Symposia

Ciba Foundation Symposium 174 ix + 316 pages, 23 figures, 6 tables

British Library Cataloguing in Publication Data A catalogue record for this book is available from the British Library

ISBN 0 471 93866 1

Phototypeset by Dobbie Typesetting Limited, Tavistock, Devon. Printed and bound in Great Britain by Biddles Ltd, Guildford.