Kenneth Williford
Department of Philosophy
The University of Iowa
Iowa City, IA 52242, U.S.A.
E-mail: kwillifo@zeus.ia.net


This volume of essays on Searle and artificial intelligence is a first-rate opportunity to re-enter, and re-explore, the Chinese Room. Now standing more than twenty years strong, John Searle’s Chinese Room argument has become one of philosophy of mind’s most venerable cites, an obligatory ‘tourist’ destination for introductory courses but no less a true philosophical monument in contemporary thought. Part philosophical polemic, part empirical computer and brain science, and part rhetorical fun house, the grand architecture of Searle’s Chinese Room argument has not changed in the twenty years it was first built: no amount of symbol manipulation is sufficient for intentionality and the mental task of semantic understanding.

Reading this collection of essays is to take a guided tour through the Chinese Room, with ample time to cover many of its additions (like the connectionist ‘Chinese Gym’) that have been constructed over the years. The volume also offers a pre-market preview of what the ‘Chinese House’ will look like in the future, with philosophical contributors laying out the blueprint of new philosophical arguments, and computer scientists designing a new generation of mental machines, some of them, according to their patrons, having already moved in. The volume, however, also succeeds in tempting the reader to imagine opening the door to another room – a door labeled the ‘Intentional Interrogation Room.’

The Intentional Interrogation Room is not part of Searle’s original design, nor explicitly described in the present volume. Nevertheless, the existence of such a room is suggested by a number of contributors (especially Chs. 4, 9, 10, 16, and
17; see also Dennett (1984), pp. 74–75). As it turns out, the door to the Intentional Interrogation Room is typically locked, for it is here that all the human Turing test participants ‘voted’ most likely to be a computer are de-briefed. After all, the Turing test affords not only the operational criteria for identifying those computers most human-like in their conversational repartee; the Turing test also identifies those humans that are most computer-like in their teletype responses. There are, as it were, two annual Turing test luminaries: the most human-like computer and the most computer-like human (see www.loebner.net for the former). The antithesis of a human-like computer, the Turing test ‘booby prize’ is awarded to the most computer-like human for having ‘fooled’ the most number of participants into thinking that they are a computer contestant rather than a human foil.

The Intentional Interrogation Room is not for the faint of heart. For it is here that you must convince a skeptical, but no less reasonable, ‘interrogator’ that you are not a ‘mere symbol manipulator,’ that you really do understand conversations about Shakespeare or Asian cooking, that you really do mean what you say, that you really do have feelings, hopes and desires – that you are not, to put it bluntly, an elaborate, albeit biologically natural, intentional hoax. Unlike the traditional Chinese Room where Searle envisions himself pretending to be a computer, the Intentional Interrogation Room requires you to prove that you are not pretending to be a ‘person’: that you are, in fact, not a ‘computer’ – not some sort of biological cyborg mindlessly, as it were, going through the paces of apparently intentional behavior. After all, Turing’s (1950) original description of the ‘imitation game’ does not envision an interviewer trying to distinguish between human-generated and computer-generated responses, but an ‘interrogator.’ Let’s listen in on how things are proceeding.

**IIR interrogator:** ... But why do you claim that – unlike the Turing test computer contestants – you actually do understand what the words mean, what the ‘semantic content’ of the words is? You say that you understand what the word ‘dog’ means. And your justification for this claim is that you know what ‘dog’ refers to. But your explanation of what ‘dog’ refers to seems to be, sorry to say, pathetically simple: ‘dog’ refers to dog. Surely you do not need a lot of your so-called distinctive mental powers to figure that one out. Wait... don’t tell me. ‘Cat’ refers to cat, right?

And then you go on to say that you know what ‘dog’ means, because you can provide a definition of ‘dog’. ‘Dog’ means something more or less like ‘domesticated canine animal’. But this too is not very convincing. After all, my electronic pocket dictionary can do that. Your so-called definition of ‘dog’ is merely the ‘output’ product of the ‘input’ question ‘What does dog mean?’ and surely you can do better than merely provide the concatenated symbols ‘domesticated canine animal’ when prompted with the symbol ‘dog’.

But then you say that you also have ‘mental imagery’ associated with the word ‘dog’. But this, like a definition for ‘dog’ is at most some sort of brute association. And that’s not particularly intelligent – brute memorization? When you say ‘dog’,