Eksamensoppgave i PSYPRO4412 Kognitiv psykologi I

Faglig kontakt under eksamen: Robert Biegler
Tlf.: 73 59 19 60

Eksamensdato: 10.12.13
Eksamens tid (fra-til): 09:00 – 13:00
Hjelpemiddelkode/Tillatte hjelpemidler: Ingen

Annen informasjon:

Målform/språk: Bokmål
Antall sider: 2
Antall sider vedlegg: 0

Kontrollert av:

____________________________  _______________________
Dato  Sign

Langsvarsoppgaver:
Besvar en av de to oppgavene
(teller 35% av den endelige karakteren)

1. Gjør rede for nevrokognitive aspekter ved smerte og subjektive somatiske symptom

2. Hva er nevrofeedback? Kan man gjennom nevrofeedback oppnå noe som ville være vanskelig eller umulig å oppnå ved bare å snakke med folk?

Kortsvarsoppgaver
Besvar to av tre oppgaver
(hvert svar teller 20 % av den endelige karakteren)

1. Gjør rede for Happé et al.s argument om at det ikke bare er en enkelt årsak til autisme, gjør også rede for implikasjonene av argumentet.

2. Diskuter hvilke interaksjoner det er mellom gener og omgivelser og hvilken betydning disse interaksjonene har for psykologien.

3. Hva er «foreward modeling» og hva kan skje hvis det går galt?

Metode spørsmål
Besvar en av de to oppgavene
(teller 25% av den endelige karakteren)

1. Evalueringen av psykologiske tester som måleinstrumenter er basert på to beslektede begrep: pålitelighet og troverdighet (reliabilitet og validitet).
Beskriv disse, hvordan de blir vurdert og forklar sammenhengen mellom dem.

2. Tester brukes ofte til å fatte viktige beslutninger om individer.
Forklar hvordan interaksjonen mellom baserate (base rate, BR)), seleksjonsrate (selection rate, SR) og test validitet (test validity) påvirker kvaliteten på disse beslutningene.
Long questions (Choose one of two, counts as 35% of the final grade)

**Question 1:** Gjør rede for nevrokognitive aspekter ved smerte og subjektive somatiske symptomer. (Discuss the neurocognitive aspects of pain and subjective somatic symptoms.)

A basic concept here is that pain is not only a function of tissue damage, but does and should depend on other factors as well. Among these are attention, interpretation and learning. Attention is directly relevant in situations when the long-term cost of injury is less than the risk posed by a greater threat. You climb a tree is a lion is after you, never mind the torn fingernails and twisted ankle. Feeling pain at such moments would only be a distraction, so it would be useful if such signals could be modulated, even down to zero. Melzack’s gate control theory postulates the dorsal horn of the spinal chord does not just passively transmit signals to the brain, but that it also receives top-down modulation that can enhance, weaken or block pain signals. The interpretation of pain signals should also modulate how much attention pain receives compared to competing information. The same pain signal merits different degrees of attention depending in whether it is interpreted as either a minor injury that will pass or as a sign of a developing serious long-term problem. That interpretation, in turn, should depend on past experience, and therefore learning. If a particular pain in the past turned out to indicate a minor injury that healed quickly, that will bias the interpretation and therefore attention in a different way than if past experience indicates a serious problem. A further effect of learning comes from it being more useful to prevent aggravating an injury rather than to start making it worse, then stop because it hurts. Therefore pain may be felt already when moving in ways that predicted imminent pain in the past. That is the general background, and students may choose to focus then on how therapy can intervene in these processes.

Students may also or instead choose to go further into the anatomical details of where pain signals are modulated, where the afferent pain signals and the efferent modulating signals come from, and what role hormones play.

A student who brings in Singer et al.’s study on the modulation of empathic pain would be making connections that have not been outlined in lectures. Briefly, Singer et al. took fMRI scans of people who either experienced electric shock at one of two levels, or believed that these shocks were administered to one of two other people. These two people were, unknown to the subjects, confederates of the experimenters, and had in an earlier game involving the division of money played either fairly or unfairly. Both men and women showed similar activation of their pain network (Melzack’s neuromatrix) when they believed that the fair player was being shocked. In men, this empathic activation of pain disappeared when they believed that the unfair player received shock. Instead, they tended to show more activation of the nucleus accumbens, whose activity generally correlates with reward. In this case, nucleus accumbens activity when the unfair player seemed to receive shock was correlated with men’s desire for revenge. The study indicates that humans can feel someone else’s pain, but that this empathic pain is modulated by how one feels about the other person, at least in men. If anyoje goes into that, it’s a bonus.
**Question 2:** Hva er neurofeedback? Kan de gjøre hva er vanskelig eller umulig å gjøre ved å snakke med folk? What is neurofeedback? Can it do anything that would be difficult or impossible to achieve by talking to people?

Eksamensoppgaven tar utgangspunkt i en labøvelse studentene har gjennomgått. 10 har gjennomført en qEEG-måling før og etter neurofeedback treninger, der resten av kullet har deltatt som observatører. I tillegg har de to artikler om neurofeedback som pensum.

Drøftinger har vært rundt hvilke mekanismer som neurofeedback berører, altså operant betinging av hjernebølger som i dette tilfelle har ført til forbedret evne til konsentrasjon. I tillegg har studentene deltatt i diskusjoner på i hvilken grad neurofeedback kan føre til bedringer om det er snakk om lidelser som ikke har et tydelig fysiologisk grunnlag, som ved traumatisering.

Det er ønskelig at besvarelsen tar opp muligheten til å endre uttrykket av hjernebølger, at det er betydningsfullt at hjernebølger kan framstå med forskjellige frekvenser og at det er vist empirisk at endring av hjernebølgeuttrykk gir funksjonelle bedringer både i normale personer og i pasienter. I tillegg bør det komme en drøfting av treningseffekter både der det er et fysiologisk grunnlag for en lidelser og der det ikke er det. Akkurat her er det få absolute svar, så de er oppfordret til å drøfte og eventuelt å spekulere.

**Short questions (Choose two of three, each answer counts as 20% of the final grade)**

**Question 1:** Gjør rede om Happé et al.’s argument at det er ikke bare en eneste årsak av autisme, og dennes implikasjoner. (Discuss Happé et al.’s argument that there is no single cause of autism and its implications.)

Happé et al. make two fundamental claims: 1) that the three core traits of autistic spectrum disorders, namely special interests and repetitive actions, social difficulties and communication difficulties, are only weakly correlated, and 2) that along each of these three dimensions the distribution of the strengths of each trait is unimodal.

What the correlations among the three traits are can only be established by studying the general population, because the autistic population is preselected to have all three traits.

The authors list five specific implications of their argument:

1) Behaviourally it would seem useful to measure the three aspects of the triad separately, rather than rely of global ratings of autism severity or ratings that focus exclusively on social functioning.

2) Molecular genetic studies should abandon the search for ‘genes for autism’ as a whole, instead focusing on genes associated with the individual traits. The authors do not mention that the search for ‘genes for autism’ will lead to overestimating the importance of genes that are linked to two or three of the traits because those genes will be overrepresented in a population selected to have all three traits. A student noticing that would be a sign of deeper thinking than is common at this level.

3) Heterogeneity in autistic spectrum disorders is not poor measurement or the complex unfolding of developmental processes, but an unavoidable consequence of variation among at least three largely independent dimensions of impairment.
4) The argument suggests that there may be many individuals with isolated impairments in one aspect of the triad who do not meet diagnostic criteria for any recognised disorder, but show difficulties of comparable severity to those with autism. [For example, Happé et al. mention that 59% of children who show social impairment show only social impairment, not the other two traits. Even if the other two traits were perfectly correlated, then for every two autistic individuals we should find another three with comparable social difficulties but who do not meet the diagnostic criteria for autism. Because the correlation between communicative difficulties and rigid/repetitive behaviour is only 0.3 to 0.4, there must be even more people with just one or two traits for every individual with all three.] A diagnostic criterion that insists on the coincidence of three largely independent traits necessarily underestimates the frequency of each of the three traits.

If the three traits are caused by different genes, are associated with different brain regions and are related to different core impairments, it is likely that they will respond to different treatments. The search for a single treatment for all difficulties is pointless if they are largely independent and distracts from finding effective treatments.

**Question 2:** Diskuter interaksjoner mellom gener og omgivelser og deres betydning for psykologien. Discuss gene-environment interactions and their significance for psychology.

The critical idea here is that it is possible to inherit how much one responds to the environment. Gabbard describes this with reference to rhesus monkeys that have been reared either by their mothers, or by peers. Rhesus monkeys have a functional polymorphism in the promoter region of the 5HTT receptor, which responds to serotonin. That region comes in a long and a short form. Every monkey has two copies of the gene, one on the chromosome inherited from the mother, the other on the chromosome inherited from the father. The study described by Gabbard compares only monkeys with two long versions of the genes versus monkeys with a long and a short version. The short/short combination is not mentioned. The monkeys with a long/long genotype are equally resistant to stressors regardless of whether they were reared by their mothers or only with peers. Those with a long/short genotype are more stressed if reared only with peers. However, monkeys placed with unusually nurturing mothers tended to rise higher in the hierarchy.

Although not in the pensum, I also showed the students the results of a paper by Crespi et al, comparing long/long, long/short and short/short genotypes in humans. The more of the short allele people had, the more did negative events bias them towards depression and suicide.

Gabbard discusses the monkey study only to suggest that early experiences can have long-lasting effects. The students must therefore work out the broader implications of gene-environment interactions for psychology by themselves. I consider one of the implications to be that any attempt to classify behaviour as being caused by either nature or nurture, as if these were mutually exclusive categories, as fundamentally wrong. Inheriting how much one responds to the environment simply doesn’t fit into that categorisation. It is as wrong to believe that genes are destiny as it is that if the environment has some influence over a trait, then that trait could not have a genetic component. Further, gene-environment interactions make it unlikely that a single treatment method will be universally effective.
**Question 3:** What is forward modelling, and what can happen if it goes wrong?

Forward modelling generates a prediction of the consequences of body movements, predicting both the trajectory of effectors (forward dynamic model) and the sensory consequences of actions (forward output model). The function of the forward output model is to separate sensory input caused by events in the environment from sensory input caused by one’s own actions. The method is to compare an efference copy (copy of motor commands) is sensory feedback, usually by subtraction of predicted sensory information from perceived information. What remains is then perceived as caused by external events. If the forward modelling goes wrong, one’s own movements may feel as if not under one’s own control. That wrong perception classifies as a hallucination. In Frith’s theory, that hallucination is one necessary condition for delusions of control. It is not a sufficient condition. The other necessary condition is a low threshold for attribution of agency. Students are not required to discuss that second condition in depth.

**Methods questions (Choose one of two, counts as 25% of the final grade)**

**Question 1:** The evaluation of psychological tests as measurement instruments centres on two related issues (reliability and validity). Describe these issues, how they are assessed, and explain their relationship.

Evaluasjon av psykologiske tester er basert på to beslektet konsepter: pålitelighet og troverdighet. Beskrive disse, hvordan er disse vurdert og forklar deres forbindelse.

Generally speaking, reliability measures how reproducible results are. Science is in the business of delivering reproducible results. More specifically, true score theory defines reliability as the proportion of total variance that is true variance, rather than error variance. The book fails to emphasise, although I do so in my lectures, that reliability thus defined is not a property of a test, but a property of the combination of test and population. It is the error variance that is a property of the test. If I test individuals who are identical with respect to the measured trait, then all obtained variance will be error variance, and the test will appear to be utterly unreliable.

The book then points out that a decision must be made what is to be considered true variance and what error variance. The example is a test of mechanical aptitude. Should noise, sleep deprivation and other stress be treated as error variance? That would be appropriate if optimal performance is to be measured. However, if an air force were interested in typical performance as a mechanic at a frontline base, response to stress should be seen as part of true variance. That is one of the links between reliability and validity, but Murphy & Davidshofer fail to point this out.

Murphy & Davidshofer then present a logical progression through four ways of measuring reliability:

Test-retest has the drawback that the first test will affect behaviour in the second test. Such reactivity, catalyst effects and carryover effects can be reduced by creating alternate forms, two different but equivalent tests. That is a more effort, and still requires two test administrations. However, a logically equivalent procedure is the split half method of dividing a test into two and treating that as two equivalent tests. There is a question how the test should be divided. Using all possible divisions and averaging the results happens to be equivalent to the internal consistency method of calculating inter-item correlations.
Measurement validity is concerned with whether we measure what we intend to measure. Content validity can be applied to very well defined domains. It is possible to establish what is and what is not arithmetic. Although there is no quantitative measure, only the opinion of experts, in a well defined domain we can expect agreement. This approach fails for ill-defined constructs such as masculinity or delusion-proneness. Then it becomes necessary to measure construct validity. That is done by estimating convergent validity (does a new test measure the same as other tests of a construct) and divergent validity (are the test scores uncorrelated with the scores of tests that measure unrelated traits). Factor analysis and the multitrait-multimethod approach are useful.

Criterion-related validity is concerned with whether a test can be used to make predictions about individuals. The statistically ideal procedure is to administer the test, then observe the outcome without any selection based on test scores. This predictive validity is discussed in the context of aptitude tests used to select employees. The practical difficulty is then that unsuitable applicants must do the job for long enough that their real life performance can be assessed. That might be expensive or even unethical in the case of safety-critical occupations such as air traffic controller. Concurrent validity tests people who do a job satisfactorily, and compares them to new applicants. The two fundamental problems are that one of the differences is experience, while in many cases the test is supposed to measure potential, and that there is range restriction. Most of the incompetent have already been kicked out of the job, so only a restricted range of traits is available for comparison.

Regarding the link between reliability and validity, the only thing that Murphy & Davidshofer mention explicitly is that a test can only be valid if it is reliable. However, deciding what is to be considered noise and what part of the measurement is also a validity issue in the measurement of reliability. The best students may notice that, with the exception of the test-retest method, there is a further validity issue in all the methods of testing reliability. If the test measures more than one factor, only items that load onto the same factor may be paired up. For example, if the reliability of the NEO-PI were measured by inter-item correlations, there is the problem that items measuring different personality dimensions are supposed to have low correlations. Items that measure one personality dimensions must be treated as one test. Thus the reliability of the NEO-PI would be assessed by dividing it into five separate tests, calculating the validity coefficient of each test, then combining those five coefficients into a single estimate. I do not expect that students write down the equation for doing that, but to get an A on this question, a student should know what the problem is.

**Question 2:** Tester brukes ofte til å fatte viktige beslutninger om individer. Forklar hvordan interaksjonen mellom baserate (base rate, BR)), seleksjonsrate (selection ratio, SR) og test validitet (test validity) påvirker kvaliteten for disse beslutningene. Tests are often used to make important decisions about individuals. Explain how the interaction between base rate (BR), selection ratio (SR), and test validity affects the quality of these decisions.

Murphy & Davidshofer’s discussion of base rate and selection rate is concerned with estimating how much a test can improve a decision, as a function of its validity coefficient. Their discussion suffers from not making explicit that their two examples use different denominators, and when it makes sense to use which denominator.

Murphy & Davidshofer examine the effects of SR and BR in the case of aptitude testing to select job applicants. One of their examples is a big table of how many qualified people you
pick depending on validity, SR and BR. There they calculate what proportion of applicants will be qualified out of those you accept. So if there are 100 applicants, you select 50, and 20 are qualified out of the 50 you accept, they say you got 40% success. That denominator makes sense for assessing whether a test improves the decision to hire job applicants. If I use a test to choose someone for a job, for example I run a big research lab and I offer four PhD positions, then I care only how many of those four turn out to be good students. I care about the proportion of true positives out of all positives, the proportion of good students out of all those whom I give a PhD position. I simply do not have the resources to offer a PhD to everyone who is good and wants one, so I do not consider the false negatives to be my responsibility. In that situation, as long as I choose someone, as long as my selection ratio is above 0, the lower the selection ratio the higher the success rate for a given validity and base rate. Also, the higher the base rate, the greater the success rate for a given validity and base rate. Greater validity always increases success.

For the other example, Murphy & Davidshofer calculate the proportion of true positives as a proportion of how many apply, rather than as a proportion of how many are selected. So if you select the same 50 out of 100 people, and 20 of them turn out to be qualified for the job, now they say that you have 20% success, instead of the previous 40%. That denominator would make sense if, for example, I want to decide who out of a number of depressed patients should get electro convulsive therapy. Because of its side effects, I do want to avoid false positives. But because depressed people have a high risk of suicide, I also want to avoid false negatives. So in this case, I may calculate my success rate relative to the whole population. Murphy & Davidshofer offer an equation that describes proportion of true positives as a function of validity, BR and SR. Providing the equation is nice, but not necessary. A qualitative description would be that if selection ratio is 0 or 1, the test can't make any difference to the decision. The closer the selection ratio is to either of these values, the less room there is to improve the decision by swapping one choice for another. The same is true for base rate. If either hardly anyone is qualified for a job, or if most are qualified, a test can do little to improve the decision.

When converting from one denominator to the other, the table and the equation should give the same result. They don't, at least not in the edition in which I looked. When I got hold of Murphy to ask, he confirmed that the two calculations conflict, which he hadn't noticed before. He took both table and equation out of the statistical literature, and because he didn't try converting results to the same denominator, the contradiction slipped through. This should be fixed in the next edition. If a student mentions that, it is a bonus, but it is not necessary to discuss this issue.